

**LSA/SCPH SEARCH: "EP 97303206.3"**

**File 351:DERWENT WPI 1963-1999/UD=9915;UP=9915;UM=9915**  
(c)1999 Derwent Info Ltd

**S1 1 AN=EP 97303206**

DIALOG(R)File 351:DERWENT WPI  
(c)1999 Derwent Info Ltd. All rts. reserv.

011575881

WPI Acc No: 97-552362/199751

XRAM Acc No: C97-176330

Isolated and crystalline herpes virus proteases - useful in assays for inhibitors, etc.

Patent Assignee: SMITHKLINE BEECHAM CORP (SMIK )

Inventor: ABDEL-MEGUID S S; CULP J; DEBOUCK C M; HOOG S S; JANSON C A; QIU X; WHITLOCK S

Number of Countries: 020 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Main IPC	Week
<b>EP 807687</b>	<b>A2</b>	<b>19971119</b>	<b>EP 97303206</b>	<b>A</b>	<b>19970512</b>	<b>C12N-015/57</b>	<b>199751 B</b>
JP 10075794 A		19980324	JP 97183392	A	19970515	C12N-015/09	199822

Priority Applications (No Type Date): US 9739191 A 19970227; US 9618616 A 19960515; US 9622470 A 19960726; US 9624416 A 19960821; US 9630901 A 19961114; US 9735973 A 19970121

Cited Patents: No-SR.Pub

Patent Details:

Patent	Kind	Lan	Pg	Filing Notes	Application	Patent
EP 807687	A2	E	494			

Designated States (Regional): AT BE CH DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE SI

JP 10075794 A 731

Abstract (Basic): EP 807687 A

A composition comprising a herpes virus protease in crystalline form (A) is new.

USE - The crystalline proteases of HSV1, HSV2, CMV and VSV are used for identifying inhibitors of these proteases, such inhibitors are useful in the treatment of herpes virus infection.

Dwg.0/45

Title Terms: ISOLATE; CRYSTAL; HERPES; VIRUS; USEFUL; ASSAY; INHIBIT  
Derwent Class: B04; D16

International Patent Class (Main): C12N-015/09; C12N-015/57

International Patent Class (Additional): A61K-038/46; A61K-045/00;  
C12N-009/50; C12N-015/09; C12R-001-92; C12R-001-19

File Segment: CPI

**LSA/SCPH SEARCH: "EP 97303206.3"**

Manual Codes (CPI/A-N): B04-L05C; B12-K04E; D05-H09

Chemical Fragment Codes (M1):

\*01\* M423 M710 M781 M903 N102 P831 Q233 R032 V802 V814

\*02\* M423 M750 M903 N102 Q233 R032 V803 V814

Chemical Fragment Codes (M6):

\*03\* M903 P831 Q233 R515 R521 R624 R633

?

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

**EP 0 807 687 A2**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:

19.11.1997 Bulletin 1997/47

(51) Int. Cl.<sup>6</sup>: **C12N 15/57**, **C12N 9/50**

(21) Application number: **97303206.3**

(22) Date of filing: **12.05.1997**

(84) Designated Contracting States:

**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE**

Designated Extension States:  
**SI**

(30) Priority: **15.05.1996 US 18616**

**26.07.1996 US 22470**

**21.08.1996 US 24416**

**14.11.1996 US 30901**

**21.01.1997 US 35973**

**27.02.1997 US 39191**

(71) Applicant:

**SMITHKLINE BEECHAM CORPORATION  
Philadelphia Pennsylvania 19103 (US)**

(72) Inventors:

• **Abdel-Meguid, Sherin Salaheldin  
King of Prussia, Pennsylvania 19406 (US)**

• **Qiu, Xiyang**

**King of Prussia, Pennsylvania 19406 (US)**

• **Whitlock Smith, Ward, Jr.**

**King of Prussia, Pennsylvania 19406 (US)**

• **Janson, Cheryl Ann**

**King of Prussia, Pennsylvania 19406 (US)**

• **Hoog, Susan S.**

**King of Prussia, Pennsylvania 19406 (US)**

• **Culp, Jeffrey**

**King of Prussia, Pennsylvania 19406 (US)**

• **Debouck, Christine Marie**

**King of Prussia, Pennsylvania 19406 (US)**

(74) Representative:

**Giddings, Peter John, Dr. et al**

**SmithKline Beecham plc**

**Corporate Intellectual Property,**

**Two New Horizons Court**

**Brentford, Middlesex TW8 9EP (GB)**

(54) **Herpesviral proteases, compositions capable of binding them and uses thereof.**

(57) Novel herpes viral protease crystalline structures are identified which have an active site formed by the three amino acids Ser, His and His. Also disclosed are methods of identifying inhibitors of these proteases and/or active sites.

**EP 0 807 687 A2**

**Description**Technical Field of the Invention

- 5 The invention relates to the identification of a novel protease catalytic active site and methods for enabling the design and selection of inhibitors of proteases, esterases, ligases and other hydrolases with that active site.

Background of the Invention

- 10 Herpesviridae is a family of envelope DNA viruses comprising three subfamilies, alpha, beta and gamma herpesviridae. The alpha subfamily includes herpes simplex virus (HSV) 1 and 2, and varicella zoster virus (VZV). The beta subfamily includes cytomegalovirus (CMV) and human herpes virus 6 (HHV-6) and human herpes virus 7 (HHV-7). The gamma subfamily includes Epstein-Barr virus (EBV) and human herpes virus (HHV-8).

- The human herpes viruses are responsible for a variety of disease states from sub-clinical infections to fatal disease states in the immunocompromised. As one example, VZV is known to cause a number of serious diseases: chickenpox, shingles and post-herpetic neuralgia [S. Straus, *Ann. Neurol.*, 35:S11-S12 (1994)]. As another example, HSV-1 is acquired in childhood when it causes a self-limiting gingivostomatitis. The virus remains latent in the dorsal root ganglia and is reactivated later in life as cold sores in about one third of the population. HSV-1 is also a cause of keratitis, resulting in more than 300,000 cases per year in the US. HSV-2 is usually acquired through sexual contact and gives rise to genital herpes. Human CMV is a ubiquitous opportunistic pathogen that can result in life threatening infections in congenitally infected infants, immunocompromised individuals and immunosuppressed cancer and transplant patients.

- Each of these members of the herpes virus families encodes a serine protease that is essential for its replication [F. Liu, & B. Roizman, *J. Virol.*, 65:5149-5156 (1991) (Liu I); F. Liu & B. Roizman, *Proc. Natl. Acad. Sci.* 89: 2076-2080 (1992) (Liu II); F. Liu & B. Roizman, *J. Virol.*, 67: 1300-1309 (1993) (Liu III); A. R. Welch *et al.*, *J. Virol.*, 67: 7360-7372 (1993) (Welch I); E. Z. Baum *et al.*, *J. Virol.*, 67: 497-506 (1993); J. T. Stevens *et al.*, *Eur. J. Biochem.*, 226: 361-367 (1994); A. R. Welch *et al.*, *J. Virol.*, 69: 341-347 (1993) (Welch II); D. L. Hall & P. L. Darke, *J. Biol. Chem.*, 270: 22697-22700 (1995); Weinheimer *et al.*, *J. Virol.*, 67: 5813-5822 (1993); M. Gao *et al.*, *J. Virol.*, 68: 3702-3712 (1994); C. L. Dilanni *et al.*, *J. Biol. Chem.*, 268: 25449-25454 (1993) (Dilanni I); C. L. Dilanni *et al.*, *J. Biol. Chem.* 269: 12672-12676 (1994) (Dilanni II); P. J. McCann III *et al.*, *J. Virol.*, 68: 526-529 (1994)]. These proteases each provide a potential target for therapeutic intervention.

- The proteases from these viruses are encoded as precursor proteins that catalyze their own cleavage to produce an N-terminal domain of approximately 28 kDa having full or increased catalytic activity. These protease domains show some degree of sequence homology - 20% to 40% identity between members of different subfamilies and as high as 90% identity within each subfamily. They show little sequence homology to any other known protein, including the absence of the conserved G-X-S/C-G-G [SEQ ID NO: 12] for chymotrypsin-like and G-T-S-M/A [SEQ ID NO: 13] for subtilisin-like proteases. The known herpes virus proteases all cleave a peptide bond between an alanine and a serine, but their substrate specificity beyond the scissile bond are different [A. Welch *et al.*, *J. Virol.* 69, 341-347 (1993)].

- Each known serine protease has its characteristic set of functional amino acid residues arranged in a particular three dimensional configuration to form an active site. Knowledge of the active site of such proteases and their three dimensional structure permits the use of methods of structure-based drug design to identify and develop inhibitors of the proteases [C. Verlinde and W. Hol, *Structure*, 2:577-587 (July 1994); I. D. Kuntz, *Science*, 257:1078-1082 (August 1992)]. Because the proteolytic activity of the herpesvirus-encoded protease plays an essential role in virus capsid maturation, inhibitors of the protease would thus inhibit infectious virus particle formation and thereby exert an antiviral action. For serine proteases of which trypsin is a prototype, the active site is formed by Ser, His and Asp [H. Neurath, *Science*, 224: 350-357 (April 1984)]. These three residues are known as the catalytic triad.

- There is a need in the art for novel protease active sites and catalytic sequences to enable identification and structure-based design of protease inhibitors, which are useful in the treatment or prophylaxis of viral diseases caused by viruses of the herpes family, as well as other diseases in which the target enzyme may share catalytic domains with those of the herpes family.

Summary of the Invention

- The present invention provides novel herpes protease crystalline forms. In one aspect, the present invention provides liganded and unliganded herpes HSV-2 protease, HSV-1 protease, CMV protease, and VZV protease crystalline forms, each of which is characterized by a three dimensional catalytic site formed by the three amino acid residues Ser, His, and His.

In another aspect, the present invention provides novel HSV-1 and HSV-2 protease compositions characterized by a three dimensional catalytic site of the seven amino acid residues, Ser 129, His 61, His 148, Ser 131, Cys 152, Arg

156, and Arg 157 of SEQ ID NOS: 3 and 4.

In still another aspect, the present invention provides a novel unliganded HSV-2 protease composition characterized by a three dimensional catalytic site of the seven amino acids identified above, and further containing amino acid residues Leu 27, Val 128, and Leu 130 of SEQ ID NO: 4, and optionally two water molecules Wat1 and Wat2 which are present in the liganded form.

In yet another aspect, the present invention provides an HSV-2 protease having an active site characterized by the coordinates selected from the group consisting of the coordinates of Figures 2 and 3, the coordinates of Figures 8 and 9, and the coordinates of Figures 11 and 12. In another aspect, the present invention provides an HSV-2 protease having an active site characterized by the coordinates selected from the group consisting of the coordinates of Figures 4 and 5, the coordinates of Figures 8 and 9, the coordinates of Figures 11 and 12, and the coordinates of Figures 14 and 15.

In yet a further aspect, the present invention provides a novel HSV-1 protease composition characterized by a three dimensional catalytic site of Ser 129, His 61, His 148, Ala 131, Cys 152, Arg 156 and Arg 157 [SEQ ID NO: 3]. In one embodiment, this HSV-1 protease has an active site characterized by the coordinates selected from the group consisting of the coordinates of Figure 6 or Figure 7, the coordinates of Figure 10, and the coordinates of Figure 16.

In still another aspect, the present invention provides a novel CMV protease composition characterized by a three dimensional catalytic site of four amino acid residues, Ser, His, His, and Asp. In one embodiment, the CMV protease active site is formed by at least the amino acids Ser 132, His 63, His 157, Asp 65, Cys 161 and Ser 134. In another embodiment, the CMV protease active site further contains at least one amino acid selected the group consisting of Arg 165 and Arg 166. Desirably, the CMV protease active site is characterized by the coordinates selected from the group consisting of Figure 17 or Figure 21, the coordinates of Figure 18 and the coordinates of Figure 20.

In yet another aspect, the present invention provides a novel CMV protease composition characterized by a three dimensional catalytic site of nine amino acid residues, Ser 132, His 63, His 157, Asp 65, Ser 134, Cys 161, Arg 165, Arg 166 and Asn 60.

In yet another aspect, the present invention provides a novel VZV protease composition characterized by a three dimensional catalytic site of four amino acid residues of SEQ ID NO: 5: Ser 120, His 52, His 139, and Lys 54. In another aspect, the VZV protease has a catalytic site which includes the four amino acids identified above and Ser 122, Cys 143, Arg 147 and Arg 148. In one embodiment, the VZV protease active site is characterized by the coordinates selected from the group consisting of Figure 22 or Figure 23, the coordinates of Figure 24 and the coordinates of Figure 26.

In another aspect, the present invention provides a heavy atom derivative of a herpes virus protease crystal, where the herpes virus protease is HSV1, HSV2, CMV, or VZV.

In a further aspect, the invention provides a method for identifying inhibitors of the compositions described above which methods involve the steps of: providing the coordinates of a protease structure of the invention to a computerized modeling system; identifying compounds which will bind to this structure; and screening the compounds or analogs derived therefrom identified for protease inhibitory bioactivity. In one embodiment of this aspect, the inhibitor binds to the dimeric interface of the protease molecule, or fragment thereof, of the invention.

In yet a further aspect, the present invention provides for an inhibitor of the catalytic activity of any composition bearing a catalytic domain described above. Desirably, the inhibitor disrupts the ability of the protease molecule to form a dimer.

Another aspect of this invention includes machine readable media encoded with data representing the coordinates of the 3D structure of a protease crystal of the invention, or of a catalytic site domain thereof.

In yet another aspect, the invention provides a computer controlled method for designing a ligand capable of binding to the active site domain of a herpes protease involving the steps of providing a model of the crystal structure of the active site domain of a herpes protease, analyzing the model to design a ligand which binds to the active site domain, and determining the effect of the ligand on the active site.

Other aspects and advantages of the present invention are described further in the following detailed description of the preferred embodiments thereof.

#### Brief Description of the Drawings

Fig. 1 provides an alignment of the amino acid sequences herpes proteases HHV-6 [SEQ ID NO: 2], CMV [SEQ ID NO: 1], EBV [SEQ ID NO: 6], HSV-1 [SEQ ID NO: 3], HSV-2 [SEQ ID NO: 4] and VZV [SEQ ID NO: 5]. Regions of  $\alpha$ -helices and  $\beta$ -strands in the HSV-2 protease structure are indicated by A1 through A7, and B1 through B7, respectively.

Fig. 2A-2F provides the coordinates of the residues of the catalytic triad and other residues and water molecules in the active site of the diisopropyl phosphate (DIP)-liganded HSV-2 protease.

Fig. 3A-3E provides the coordinates of the residues of the catalytic triad and other residues and water molecules in the active site of unliganded HSV-2 protease.

Fig. 4A-4MMM provides the protein coordinates of the DIP-liganded HSV-2 protease crystalline structure of the

invention. Figure 2A-2F is included within Figure 4A-4MMM.

Fig 5A-5DDD provides the protein coordinates of the unliganded HSV-2 protease crystalline structure. Figure 3A-3E is included in Figure 5A-5DDD.

Fig. 6A-6B provides the coordinates of residues in the active site region of HSV-1 protease.

Fig. 7A-7DDD provides the protein coordinates of the HSV-1 protease crystalline structure of the invention. Figure 6A-6B is included within Fig. 7A-7DDD.

Fig. 8A-8X provides the distances (in Angstrom) between any protein residue that has an atom within a 5.5 Å radius of any atom of Ser 129 covalently bonded to the DIP in the DIP-liganded HSV-2 protease [SEQ ID NO: 4].

Fig. 9A-9I provides the distances (in Angstroms) between every two atoms that are within 5.0 Å of the active site residues of the unliganded HSV-2 protease.

Fig. 10A-10B provides the distances (in Angstroms) between every two atoms that are less than 5 Å from the active site of HSV-1 protease.

Fig. 11A-11LLL provides the bond angles (in degrees) between any atom of HSV-2/DIP-modified Ser 129 and any two protein residue atoms that are within a 5.5 Å radius of the DIP modified Ser 129 [SEQ ID NO: 4].

Fig. 12A-12Q provides the bond angles (in degrees) between interresidue atoms in the active site region near Ser 129, His 61 and His 148 of the unliganded HSV-2 protease [SEQ ID NO: 4].

Fig. 13A-13D provides the bond angles between interresidue atoms in the active site region near Ser 129, His 61 and His 148 of the HSV-1 protease of this invention [SEQ ID NO: 3].

Fig. 14 provides the dihedral angles of the active site formed by Ser 129, His 61 and His 148 of the DIP-liganded HSV-2 protease [SEQ ID NO: 4].

Fig. 15 provides the dihedral angles of the active site formed by Ser 129, His 61 and His 148 of the unliganded HSV-2 protease [SEQ ID NO: 4].

Fig. 16 provides the dihedral angles of the active site formed by Ser 129, His 61 and His 148 of the HSV-1 protease [SEQ ID NO: 3].

Fig. 17A-17E provide the protein coordinates near the active site region (at amino acid residues Ser 132, His63, Ser 134, Cys161, Arg165 and His157, and including Arg166) of the CMV protease [SEQ ID NO: 1] according to this invention.

Fig. 18A-18C provide the distances in Angstroms between every two atoms that are less than 5Å from the active site for CMV protease.

Fig. 19A-19D illustrate the bond angles between interresidue atoms that are within four Å apart in the active site region near Ser132, His63, His157 and Asp65 of the CMV protease [SEQ ID NO: 1] according to this invention.

Fig. 20 provides the dihedral angles of the tetrad active site formed by Ser132, His63, His157 and Asp65 of the CMV protease [SEQ ID NO: 1].

Fig. 21A-21DD provides the protein coordinates of the CMV protease crystalline structure of the invention. Fig. 17A-17E is included within Fig. 21A-21DD.

Fig. 22A-22C provide the protein coordinates near the active site region (at amino acid residues Ser 120, Ser 122, His 52, Lys 54, Cys 143, Arg 147 and Arg 148 of SEQ ID NO: 5) of the VZV protease according to this invention. The data is reported in Protein Data Bank (PDB) format as in Figs. 3A-3E.

Fig. 23A-23SS provide the protein coordinates of the VZV protease crystalline structure of the invention, including the active site of Figure 22A-22C.

Fig. 24A-24C provide the distances in Angstroms between every two atoms that are less than 5Å from the active site for VZV protease.

Fig. 25A-25D illustrate the bond angles between interresidue atoms that are within four Å apart in the active site region near Ser 120, His 52 and His 139 of the VZV protease [SEQ ID NO: 5] according to this invention.

Fig. 26 provides the dihedral angles of the tetrad active site formed by Ser 120, His 52 and His 139 of the VZV protease [SEQ ID NO: 5].

Fig. 27A is a three dimensional ribbon diagram of the DIP-liganded HSV-2 protease dimer. The ligand diisopropyl phosphate (DIP) is shown in the active site of each monomer rendered in space filling models. The amino terminus is indicated by N. The drawing was produced using the program RIBBONS [Carson, M. J. *Mol. Graphics* 5, 103-106 (1987)].

Fig. 27B is the same structure as in Fig. 27A. viewing from 90° away.

Fig. 27C is the unliganded dimer. The structure is essentially identical to that of the DIP-liganded HSV-2 protease structure.

Fig. 28A is a three dimensional ribbon diagram of the HSV-1 protease dimer. The amino terminus is indicated by N. The drawing was produced using the program RIBBONS [Carson, M. J. *Mol. Graphics* 5, 103-106 (1987)].

Fig. 28B is the same structure as in Fig. 28A, viewing from 90° away.

Fig. 29A is a three dimensional diagram of the HSV-2 protease monomer. The ligand DIP is shown in the active site rendered as a space filling model. The amino terminus is indicated by N. The drawing was produced using the program RIBBONS [Carson, M. J. *Mol. Graphics* 5, 103-106 (1987)].

Fig. 29B is the same structure as in Fig. 29A, viewing from 90° away.

Fig. 29C is the unliganded HSV-2 monomer. The structure is the same as the DIP-liganded HSV-2 protease structure.

Fig. 30A is a three dimensional diagram of the HSV-1 protease monomer. The amino terminus is indicated by N. The drawing was produced using the program RIBBONS [Carson, M. *J. Mol. Graphics* 5, 103-106 (1987)].

Fig. 30B is the same structure as in Fig 30A, viewing from 90° away.

Fig. 31 is a topology diagram of the HSV-2 monomer of Fig. 29A and HSV-1 monomer of Fig. 30A with helices (A1 through A7) represented as cylinders, strands (B1 through B7) represented as arrows and termini as N or C. Strands B5 and B7 are next to each other. Amino acid positions are indicated.

Fig. 32A is a three dimensional diagram of the structure of CMV protease with the core  $\beta$ -barrel highlighted. The amino and carboxyl-termini are indicated by N and C. Disordered portions of the structure are represented by dashed lines. The diagram was drawn with the program MOLSCRIPT [P. Kraulis, *J. Appl. Crystallogr.*, 24: 946-950 (1991)].

Fig. 32B is the same structure as in Fig. 32A, viewing from 90° away.

Fig. 33A is a three dimensional diagram of the structure of VZV protease with the core  $\beta$ -barrel highlighted in two shades of gray. The amino and carboxyl-termini are indicated by N and C. A disordered portion of the structure is represented by a dashed line. The diagram was drawn with the MOLSCRIPT program [P. Kraulis, *J. Appl. Crystallogr.*, 24:946-950 (1991)].

Fig. 33B is the same structure as in Fig. 33A, viewing from 90° away.

Fig. 34 is a topology diagram of the CMV monomer of Fig. 32A with helices (A1 through A7) represented as cylinders, strands (B1 through B7) represented as arrows and termini as N or C. Strands B5 and B7 are next to each other. Amino acid positions are indicated.

Fig. 35 is a topology diagram of the VZV monomer of Fig. 33A with helices (AA through A7) represented as cylinders, strands (B1 through B7) represented as arrows and termini as N or C. Strands B5 and B7 are next to each other. Amino acid positions are indicated.

Fig. 36A is a three dimensional diagram of the CMV protease dimer, viewing perpendicular to the two-fold axis, which was drawn using the MOLSCRIPT program.

Fig. 36B is the dimer of Fig. 36A, viewing parallel to the two-fold axis. The two parallel helices are indicated by A6. The active site regions are represented by Ser at the Ser132 positions [SEQ ID NO: 1].

Fig. 37 is a three dimensional diagram of the VZV protease dimer, viewing parallel to the two-fold axis, which was drawn using the MOLSCRIPT program, with each subunit in a different shade of grey.

Fig. 38A is a model of the DIP-liganded HSV-2 protease active site in a thick stick representation. All carbon atoms are in light shading; nitrogen, oxygen and sulfur atoms in dark shading. The catalytic residues are Ser 129, His 61 and His 148. Other residues of importance are Arg 156, Arg 157 and possibly Ser 131, Cys 152, Leu 27, Val 128, Leu 130 [SEQ ID NO: 4] and the two water molecules Wat1 and Wat2 shown as spheres. Hydrogen bonds between the residues in this active site are shown with dotted lines.

Fig. 38B is a superposition of the unliganded HSV-2 protease (dark) and the DIP-liganded HSV-2 protease (light) structures where the DIP ligand has been removed from the active site serine for clarity.

Fig. 38C is the superposition between the active site of DIP-liganded HSV-2 protease and that of the classical serine protease  $\gamma$ -chymotrypsin complexed with mono isopropyl phosphate (MIP). In light shading is the HSV-2 protease in the identical orientation as in Fig. 38A, and in dark shading are the  $\gamma$ -chymotrypsin active site residues. Labels are those of protease  $\gamma$ -chymotrypsin. Hydrogen bonding in the oxyanion hole of  $\gamma$ -chymotrypsin is shown.

Fig. 38D is the superposition of DIP-liganded HSV-2 protease with CMV protease, illustrating the similarities and differences between the active sites of these two enzymes. Hydrogen bonding is shown for CMV protease but labels are for HSV-2 protease.

Fig. 39A is the HSV-1 protease active site in a thick stick representation. All carbon atoms are in light shading; nitrogen, oxygen and sulfur atoms in dark shading. The catalytic residues are Ser129, His61 and His148 [SEQ ID NO: 3]. Other residues of importance are Arg 156, Arg 157 and possibly Ala 131 and Cys 152 [SEQ ID NO: 3]. Hydrogen bonds between the residues in this active site are shown with dotted lines.

Fig. 39B is the superposition between the active site of HSV-1 protease and that of the classical serine protease trypsin. In light shading is the HSV-1 protease in the identical orientation as in Fig. 39A, and in dark shading are the trypsin active site residues. Labels are shown for both the active site of trypsin (Ser 195, His 57, Asp 102) and HSV-1 protease (Ser 129, His 61, and His 148).

Fig. 39C is the superposition of HSV-1 protease with HSV-2 protease, illustrating the similarities and differences between the active sites of these two enzymes. In light shading is the HSV-1 protease in the identical orientation as in Fig. 39A, and in dark shading are the HSV-2 protease active site residues (numbering of the residues is identical between HSV-1 protease and HSV-2 protease).

Fig. 39D is the superposition of HSV-1 protease with CMV protease, illustrating the similarities and differences between the active sites of these two enzymes. In light shading is the HSV-1 protease in the identical orientation as in Fig. 39A, and in dark shading are the CMV protease active site residues.

Fig. 40A is a drawing of the VZV protease active site in ball-and-stick representation drawn with MOLSCRIPT program. All carbon atoms are in light shading; nitrogen, oxygen and sulfur atoms in dark shading. The catalytic residues are Ser 120, His 52 and His 139 of SEQ ID NO: 5. Other residues of importance are Arg 147, Arg 148 and possibly Ser 122 and Cys 143 of SEQ ID NO: 5. The postulated proton transfer pathway is shown by dashed lines.

Fig. 40B is the superposition between the active site of VZV protease and that of the classical serine protease trypsin. In light shading is the VZV protease in the identical orientation as in Fig. 40A, and in dark shading are the trypsin active site residues. Labels are those of trypsin. The proton transfer pathway in trypsin are depicted in dotted lines.

Fig. 41A is a drawing of the stereoview of the CMV protease active site in ball-and-stick representation drawn with MOLSCRIPT program. The catalytic residues are Ser132, His63, His157 and optionally Asp65 [SEQ ID NO: 1]. Other residues of importance are Arg 165, Arg 166 and possibly Ser 134 and Cys 161 [SEQ ID NO: 1]. The postulated proton transfer pathway is shown by dashed lines. The side chain of Arg165 is disordered in the structure and hence omitted in the drawing. H<sub>2</sub>O is an ordered water molecule.

Fig. 41B is a stereoview of the superposition between the active site of CMV protease and that of the classical serine protease trypsin. The CMV protease is in the identical orientation as in Fig. 41A. Labels are those of trypsin. The proton transfer pathway in trypsin are depicted in dotted lines.

Fig. 42 provides the statistics of structure determination for HSV-2 protease.

Fig. 43 provides the data collection statistics for native and heavy atom derivatives of CMV protease.

Fig. 44 is a stereoview of the superposition between the Ca trace of VZV protease and that of the CMV protease. The light color strand is the CMV protease and the dark color strand is the VZV protease. In spite of the limited sequence homology and apparent conformational differences in some helical or loop regions, the core  $\beta$ -barrel of the VZV protease superimposes quite well with that of the CMV protease. Excluding regions that are drastically different ( $> 4 \text{ \AA}$ ), the root-mean-square (rms) difference is only  $1.3 \text{ \AA}$  between 142 (60%) Ca atoms from VZV and CMV.

Fig. 45A is the complete sequence of the VZV construct H6(N)VZV [SEQ ID NO: 7] containing an authentic protease domain preceded at the amino-terminus by six histidine residues (underlined) followed by an enterokinase cleavage site (bold, underlined).

Fig. 45B is the complete sequence of the VZV construct LQA-H6(C) VZV [SEQ ID NO: 8] containing an authentic protease domain followed by six histidine residues (underlined).

Fig. 45C is the complete sequence of the VZV construct LQAS-H6(C) VZV [SEQ ID NO: 9] containing an authentic protease domain followed by a serine residue and six histidine residues (underlined).

Fig. 45D is the complete sequence of the VZV construct LQAS-12aa ext H6(C) VZV [SEQ ID NO: 10] containing an authentic protease domain followed by a serine residue, 12 residues normally found after the LQAS R-site (bold underlined) and six histidine residues (underlined).

Fig. 45E is the complete sequence of the VZV construct  $\Delta 9$  LQAS-12 aa ext H6(C) VZV [SEQ ID NO: 11] containing an authentic protease domain deleted at the amino-terminus (first nine natural residues removed and Cys 10 replaced by Met) and followed at the carboxyl-terminus by a serine residue, 12 residues normally found after the LQAS R-site (bold underlined) and six histidine residues (underlined).

#### Detailed Description of the Invention

The present invention provides novel herpes virus family protease crystalline structures, novel herpes virus protease active sites, and methods of use of the crystalline forms and active sites to identify protease inhibitor compounds (peptide, peptidomimetic or synthetic compositions) characterized by the ability to inhibit binding to the active site of herpes proteases. The herpes virus protease compositions of the invention are characterized by a three dimensional active catalytic site of conserved amino acid residues, Ser, His, and His.

For HSV-1 and HSV-2 proteases, these residues are located at aa Ser 129, His 61 and His 148 of SEQ ID NO: 3 and 4, respectively. For CMV protease, these residues are located at aa Ser 132, His 63 and His 157 of SEQ ID NO: 1. For VZV protease, these residues are located at aa Ser 120, His 52, and His 139 of SEQ ID NO: 5.

The present invention further provides a novel HSV-2 protease composition characterized by a three dimensional active catalytic site of three conserved amino acid residues, Ser 129, His 61, His 148, seven additional amino acid residues Ser 131, Cys 152, Arg 156, Arg 157, Leu 27, Val 128, Leu 130, and two water molecules Wat1 and Wat2 (present in only the DIP-liganded HSV-2 protease structure) as defined by position in Figure 1 herein [SEQ ID NO: 4].

Also provided is a novel HSV-1 protease composition characterized by a three dimensional active catalytic site of three conserved amino acid residues, Ser 129, His 61, His 148, four additional amino acid residues Ala 131, Cys 152, Arg 156, and Arg 157, as defined by position in Figure 1 herein [SEQ ID NO: 3].

In yet another aspect, the present invention provides a novel CMV protease composition characterized by a three dimensional active catalytic site of six conserved amino acid residues, Ser 132, His 63, His 157, Asp 65, Cys 161 and Ser 134. Additionally, the CMV catalytic site may also contain either or both Arg 165 and Arg 166, as defined by their position in Figure 1 herein [SEQ ID NO: 1]. The CMV structure further reveals a novel active site tetrad, formed by

Ser132-His63-His157-Asp65 [SEQ ID NO: 1]. In still another aspect, the present invention provides a novel CMV protease composition characterized by a three dimensional active catalytic site of six conserved amino acid residues, Ser 132, His 63, His 157, Cys 161, Arg 165, and Arg 166 and three non-conserved amino acid residues Ser 134, Asp 65, and Asn 60, as defined by position in Figure 1 herein [SEQ ID NO: 1].

5 The invention further provides a novel VZV protease composition characterized by a three dimensional active catalytic site of three conserved amino acid residues, Ser 120, His 52, His 139, and five additional amino acid residues Ser 122, Cys 143, Arg 147, Arg 148 and Lys 54, as defined by the amino acid positions in Fig. 1 [SEQ ID NO: 5].

### 1. The Novel Protease Crystalline Three-Dimensional Structure

10 The present invention provides novel protease crystalline structures based on the herpes proteases. The three dimensional (3D) structure of the HSV-1, HSV-2, CMV and VZV proteases provided herein reveal a unique fold that has not been reported for any serine protease, and a novel active site consisting of a novel catalytic triad formed by 3D interactions of the amino acids Ser, His and His. An unusual dimer interface that is important to protease activity was also found in the crystals for each of these proteases. In yet another aspect, the present invention provides for a novel herpes protease composition characterized by a dimer interface of two herpes protease molecules. Inhibition of this dimer interface by inhibitors which perturb interaction with these dimer interfaces. Inhibitors that perturb or interact with these dimer interfaces are yet another therapeutic target for the design and selection of therapeutic agents against herpes proteases.

20 Listed in Fig. 1 are known amino acid sequences for the herpes family proteases HHV-6, HSV-1, HSV-2, VZV, EBV and CMV, aligned to illustrate the homologies between them [SEQ ID NO: 1-6]. As seen in Fig. 1, when compared to members of the alpha subfamily, HSV-2, HSV-1, and VZV protease amino acid sequences are rather conserved (HSV-1 protease is 50% identical to VZV protease and 90% identical to HSV-2 protease; VZV protease is 26% identical to CMV protease); CMV protease differs in having a shorter N-terminus and two multi-residue insertions (at CMV protease amino acid residues 40-47 and 147-152, SEQ ID NO: 1).

25 According to the present invention, the crystal structure of human HSV-1, liganded and unliganded HSV-2, CMV and VZV proteases have been determined. These protease crystal structures reveal a fold and an active site that are distinct from other known, non-herpes serine proteases. Details of structure determination and refinement are presented in Figures 2 - 26 below.

30 Further refinement of the atomic coordinates will change the numbers in Figures 2 - 26, refinement of the crystal structure from another crystal form will result in a new set of coordinates, determination of the crystal structure of another herpes protease will also result in a different set of numbers for coordinates in these figures. However, distances and angles will remain the same within experimental error, and relative conformation of residues in the active site will remain the same within experimental error. Also for example, the amino acid sequence of the herpes proteases can be varied by mutation derivatization or by use of a different source of the protein, as described herein.

#### A. HSV-1 and HSV-2

40 The crystal structures of HSV-1 and HSV-2 have been determined, as described herein, and are discussed in tandem below.

The crystal structure of human HSV-1 protease has been determined at 3.5 Å resolution. The structure was determined using the method of molecular replacement (MR) and refined to an R-factor of 36.9% (10.0-3.5 Å, using  $|F_o| > 2s |F_c|$  data).

45 The crystal structures of human DIP-liganded HSV-2 protease and the unliganded HSV-2 protease have been determined at 2.5 Å and 2.8 Å resolution, respectively. The DIP-liganded HSV-2 protease structure was determined using the methods of multiple isomorphous replacement (MIR) and MR and refined to an R-factor of 20.5% (10.0-2.5 Å, using  $|F_o| > 2s |F_c|$  data) with root-mean-square deviations on bond lengths and bond angles of 0.016 Å and 1.9°, respectively. The unliganded structure was determined using difference Fourier methods using the DIP-liganded HSV-2 protease structure and refined to an R factor of 22.4 %. The root-mean-square deviations of bond lengths and bond angles are 0.017 Å and 2.1 Å.

50 Although human HSV-1 and HSV-2 proteases contain 247 amino acids, the models of these proteases are represented by 214 residues for HSV-1 protease and 217 residues for DIP-liganded HSV-2 protease. With respect to HSV-1 protease, residues 1-14 and residues in two surface loops 102-110, and 134-143 are disordered in the crystal [SEQ ID NO: 3]. With respect to HSV-2, residues 1-16 and residues in two surface loops 104-110, and 134-140 are disordered in the crystal [SEQ ID NO: 4]. The model of unliganded HSV-2 protease has 215 residues where 1-16, 104-112, and 134-140 are disordered in the crystal [SEQ ID NO: 4].

55 The fold in each of the HSV-1 and HSV-2 proteases is characterized by 7  $\beta$ -strands and 7  $\alpha$ -helices as depicted in Figures 1, 29, and 30. As discussed herein, the loop containing residues 25-55 in the CMV protease [SEQ ID NO: 1] is disordered in the crystal, but the corresponding loop in the HSV-1 and HSV-2 protease crystals are observed.

Fig. 2A-2F disclose the coordinates of the residues of the novel HSV-2 catalytic triad and other residues and water molecules within 5.5 Å of the DIP modified Ser 129 (including the DIP ligand) in the active site. These include amino acid residues Ser 129, His 61, His 148, Ser 131, Cys 152, Arg 156, Arg 157, Leu 27, Val 128, Leu 130, and, water molecules Wat1 and Wat2 of the DIP-liganded HSV-2 protease [SEQ ID NO: 4] according to this invention. Fig. 3A-3E disclose the coordinates of residues of the novel catalytic triad and other residues in the active site of the unliganded HSV-2 protease. These include residues Ser 129, His 61, His 148, Ser 131, Cys 152, Arg 156, Arg 157, Leu 27, Val 128, and Leu 130. These data are reported for crystals with lattice constants of  $A=71.7$  Å,  $B=87.4$  Å,  $C=77.3$  Å,  $\alpha=90$ ,  $\beta=90$  and  $\gamma=90$ , with a space group =  $P2_12_12_1$ . The liganded and unliganded crystal forms have the same cell dimensions and space group. The data is reported in Protein Data Bank (PDB) format, illustrating the atom, i.e., nitrogen, oxygen, carbon (at  $\alpha$ ,  $\beta$ ,  $\delta$  or  $\gamma$  positions in the atom); the amino acid residue in which the atom is located with amino acid number, and the coordinates X, Y and Z in Angstroms (Å) within the crystal lattice. Also illustrated is the occupancy of the atom, noting that each atom in the active site has a unique position in the crystal. The data also report the B or Temperature Factor, which indicates the degree of thermal motion of the atom in volume measurements (Å<sup>2</sup>). Figures 4A-4MMM and 5A-5DDD disclose the protein coordinates of the DIP-liganded HSV-2 and unliganded protease crystalline structure of the invention respectively, including the active site.

Figure 6A-6B disclose the coordinates of residues in the active site region (at amino acid residues Ser 129, His 61, His 148, Ala 131, Cys 152, Arg 156, Arg 157) of the HSV-1 protease [SEQ ID NO: 3] according to this invention. These data are reported for crystals with lattice constants of  $a=79.62$  Å,  $b=81.18$  Å,  $c=93.36$  Å,  $\alpha=115.49^\circ$ ,  $\beta=98.36^\circ$  and  $\gamma=109.18^\circ$ , with a space group =  $P1$ . The data is reported as described above for HSV-2 protease. Figure 7A-7DDD illustrate the orthogonal three dimensional coordinates in Angstroms and B factors for HSV1 protease.

Figure 8A-8X provide the distances (in Angstroms) between any protein residue that has an atom within a 5.5 Å radius of any atom of Ser 129 [SEQ ID NO: 4] covalently bonded to DIP in the DIP-liganded HSV-2 protease structure. Figure 9A-9I provide distances (in Angstroms) between every two atoms that are within 5.0 Å of the active site residues of the unliganded HSV-2 protease. The atoms are indicated in this figure by the amino acid position number in which the atom appears, followed by the atom designation, i.e., nitrogen, oxygen, etc. as described for Figures 2 and 3.

Figure 10A-10B provide the distances (in Angstroms) between every two atoms that are less than 5 Å from the active site of HSV-1 protease. The atoms are indicated as described above for HSV-2 protease. Figure 11A-11LLL provide the bond angles (in degrees) between any atom of HSV-2/DIP-modified Ser 129 and any two protein residue atoms that are within a 5.5 Å radius of the DIP modified Ser 129 (Overall Active Site Residues Only). Figure 12A-12Q provide the bond angles (in degrees) between interresidue atoms that are within 5 Ångstroms apart in the active site region near Ser 129, His 61 and His 148 [SEQ ID NO: 4] of the unliganded HSV-2 protease. Figure 13A-13D provide the bond angles between interresidue atoms that are within 5 Ångstroms apart in the active site region near Ser 129, His 61 and His 148 of the HSV-1 protease [SEQ ID NO: 3] of this invention. Figures 14 and 15 provide the dihedral angles of the active site formed by Ser 129, His 61 and His 148 [SEQ ID NO: 4] of the protease, for the DIP-liganded HSV-2 and unliganded proteases, respectively. Figure 16 provides the dihedral angles of the active site formed by Ser 129, His 61 and His 148 of the HSV-1 protease [SEQ ID NO: 3].

The novel folds of the HSV-2 and HSV-1 protease crystal structures are discussed in part D. below; the novel active sites are discussed in part E.

#### 40 B. CMV Protease

The crystal structure of human CMV protease has been determined at 2.5 Å resolution. As described in more detail in Example 3 below, a CMV protease sequence (CMV A143V) was employed in which the alanine residue at position 143 was replaced by a valine. This mutation had no effect on the enzyme activity, but was necessary to eliminate the nick after Ala 143 seen in preparations of the native enzyme [Welch I, and Baum *et al.*, cited above]. The structure was determined using MIR, and refined to an R-factor of 18.5% (7.0-2.5 Å, using  $|Fo| > 1\sigma |Fo|$  data) with root-mean-square deviations on bond lengths and bond angles of 0.017 Å and 2.2°, respectively.

Although human CMV protease contains 256 amino acids, the model of the enzyme provided herein is represented by 202 residues. Residues 1-8 and residues in three surface loops (residues 25-55, 143-153 and 205-208 of SEQ ID NO: 1) are found disordered and are not present in Figure 17A-17E. Details of structure determination and refinement are presented in Figures 17 through 21.

Figure 17A-17E disclose the protein coordinates near the active site region (at amino acid residues Ser132, His63, Ser 134, Cys161, Arg165 and His157, and including Arg166 of SEQ ID NO: 1) of the CMV protease according to this invention. These data are reported for crystal lattice constants of  $a=58.7$  Å,  $b=58.7$  Å,  $c=131.0$  Å,  $\alpha=90$ ,  $\beta=90$  and  $\gamma=90$ , with a space group =  $P4_32_1$ . The data is reported in Protein Data Bank (PDB) format, as described above for HSV-2 and HSV-1. Figure 18A-18C provide the distances in Angstroms between every two atoms that are less than 5 Å from the active site for CMV protease. Figure 19A-19D illustrate the bond angles between interresidue atoms that are within four Å apart in the active site region near Ser132, His63, His157 and Asp65 of the CMV protease [SEQ ID NO: 1] according to this invention. Figure 20 provides the dihedral angles of the tetrad active site formed by Ser132, His63,

His157 and Asp65 of the protease [SEQ ID NO: 1]. Figure 21A-21DD disclose the protein coordinates of the CMV protease crystalline structure of the invention, including the active site. See D. below for discussion of the CMV crystal structure folds and part E. for further discussion of the active site.

### 5 C. VZV Protease

The crystal structure of human VZV protease has been determined at 3.0 Å resolution. As described in more detail in the Example 6 below, a VZV protease sequence was employed in which the mutant has the N-terminal nine (9) amino acids deleted. This mutation had little effect on the enzyme activity. The structure was determined using the methods of MR and single isomorphous replacement (SIR) and refined to an R-factor of 22.3% (7.0-3.0 Å, using  $|F_o| > 1s |F_o|$  data) with root-mean-square deviations on bond lengths and bond angles of 0.014 Å and 2.1°, respectively.

The model of the VZV enzyme used in the crystal structure study consisted of residues 11-236 of SEQ ID NO: 5. The VZV structure model consists of 211 amino acids. A surface loop of aa 127-136 is disordered in the crystal, as are the last 5 amino acids at the C-terminal of VZV. In VZV, an additional helix has been observed in the region between residues 31-39 of SEQ ID NO: 5. The fold is characterized by 7β strand and 8 α-helices as depicted in Figs. 1 and 35. The loop containing residues 22-55 in the CMV protease is disordered in the crystal, but as for HSV-1 and HSV-2 proteases, the corresponding loop in the VZV protease crystal is observed. This loop is situated near the active site and is proposed to enclose part of the substrate binding groove, the S subsites [I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.*, 27:157-162, (1967)].

Fig. 22A-22C disclose only the protein coordinates near the active site region (at amino acid residues Ser 120, Ser 122, His 52, Lys 54, Cys 143, Arg 147 and Arg 148 of SEQ ID NO: 5) of the VZV protease according to this invention. These data are reported for crystals with lattice constants of  $a=90.0$  Å,  $b=90.0$  Å,  $c=117.4$  Å,  $\alpha=90^\circ$ ,  $\beta=90^\circ$  and  $\gamma=90^\circ$ , with a space group =  $P6_422$ . Fig. 23A-23SS disclose protein coordinates of the VZV protease crystalline structure. Fig. 24A-24C provide the distances in Angstroms between every two atoms that are less than 5 Å from the active site for VZV protease. Fig. 25A-25D illustrate the bond angles between interresidue atoms that are within four Å apart in the active site region near Ser 120, His 52 and His 139 of the VZV protease [SEQ ID NO: 5] according to this invention. Fig. 26 provides the dihedral angles of the tetrad active site formed by Ser 120, His 52 and His 139 of the protease [SEQ ID NO: 5].

The VZV protease crystal structure novel fold is discussed in more detail in D. below. The Novel active site is discussed in part E.

### D. The novel fold

#### 1. HSV-2 and HSV-1 Proteases

With reference to Figures 27A, B and C, the structures of the liganded and unliganded forms of HSV-2 are nearly identical. The binding of DIP to the HSV-2 protease structure does not alter the conformation of the enzyme. The root mean square deviation between the liganded and unliganded forms is 0.4 Å. The presence of the DIP changes interactions only within the active site of the enzyme but not the location of the active site or overall three dimensional structure of the enzyme.

The overall folds of the HSV-1 and HSV-2 protease dimers are comprised of two β-barrels. The folds of the HSV-2 protease are discussed above. The overall fold of the HSV-1 protease dimer is illustrated in Figs. 28A and 28B. For each protease, seven b-strands form the core of each barrel (Fig 29A,B,C for HSV-2 and 30A, 30B for HSV-1), each of which can be classified as an orthogonally packed β-barrel [(Chothia and Janin *Biochemistry*, 21: 3955-3965 (1982)], with the following exceptions: strands B6 and B7 are parallel, unlike most orthogonally packed b-barrels. Also, strand B3 (aa 65-77) is a β-bend that closes one corner of the barrel, but the other corner lacks this kind of classical closure and is maintained by only two hydrogen bonds between strand B5 (aa 127-133) and B7 (aa 161-166) [SEQ ID NO: 3 and 4]. Among well known serine proteases, the N-terminal domain of trypsin is also an orthogonally packed b-barrel, but superposition with the HSV-1 or HSV-2 protease barrel does not reveal any similarities and positions the active sites in different regions of the fold. Moreover, the β-strands (Fig. 31 for HSV-2 and HSV-1) are arranged differently in the two structures. For example, the first four strands of HSV-2 protease (B1, B2, B3 and B4) form a typical Greek Key motif while those in trypsin do not. Therefore, it is reasonable to conclude that the HSV-2 and HSV-1 protease barrels are evolutionally unrelated to other known non-herpes serine proteases.

The seven alpha helices of liganded and unliganded HSV-2 protease and HSV-1 do not surround the barrel but rather cluster towards the ends. Alpha helix A1 seals one end while helices A2, A3, A6 and A7 close the other. Of these, helices A6 and A2 with the corresponding helices of the dimer mate define a unique dimer interface that reveals an approximately 30 degree twist between the corresponding monomers. While four of the seven helices are at either end of the barrel the other two are on the same side of the structure, away from the active site.

For the DIP-liganded HSV-2 protease structure, the transition state analog inhibitor diisopropyl fluorophosphate

was added to the enzyme in which the highly reactive (P-F) linkage undergoes displacement by the serine hydroxyl group. Once bound to the nucleophile (enzyme) the ligand is referred to as diisopropyl phosphate. The DIP-liganded HSV-2 protease has the inhibitor diisopropyl phosphate covalently bound to the active site serine (Ser 129).

Unlike trypsin, the active sites of HSV-1 and HSV-2 proteases do not lie at the intersection of the two domains but rather bind in a cleft formed by one side of the barrel at strand B5 and a small loop between strands B6 and B7 made of residues 153-157 (Figs. 27A,B,C and 29A,B,C for HSV-2 and Figs. 28A, 28B and 30A, 30B for HSV-1). The DIP C $\beta$  atom of the DIP-liganded HSV-2 protease is 17 Å from the closest point between the two monomers at helix A6. In each of the HSV-1 and HSV-2 protease structures, residues 34-38 [SEQ ID NO: 3 and 4] in the large loop between B1 and A1 approach the top of the cleft but do not completely cover it. The loop of residues 134-140 [SEQ ID NO: 3 and 4] is disordered in the HSV-1 and HSV-2 structures. This loop is a region of low homology among the herpes proteases (Fig 1), and may suggest different conformations of the loop. Residues 104-110 and 104-112 are disordered in the DIP-liganded HSV-2 and unliganded HSV-2 protease structures, respectively [SEQ ID NO: 4]. Residues 102-110 of HSV-1 [SEQ ID NO: 3] are disordered in the structure. These regions represent regions of low homology among the herpes proteases; however, the corresponding regions are ordered in both the VZV and CMV proteases, where a small alpha helix is seen.

An intriguing non-crystallographic dimer interface in each of the HSV-1 and HSV-2 proteases is made up of the interactions between helices A6 and A2 with the corresponding helices in the dimer mate. The dimer interface is identical between the liganded and unliganded HSV-2 proteases. These same helices also interact in CMV but with a different orientation in relation to each other. In CMV protease the two A6 helices of each monomer are nearly co-axial. In the HSV-1 and HSV-2 proteases they are separated at the N-terminal ends by about 30 degrees, at a distance of nearly 14 Å. Because of the twist between the A6 helices in the HSV-1 and HSV-2 proteases, contacts can only be formed between the C-terminal ends of the helices. In both the liganded and unliganded HSV-2 proteases, a hydrogen bond (3.0 Å) is seen between the side chains of His 211 and Glu 207 [SEQ ID NO: 4]. When comparing the dimer interface of HSV-1, HSV-2 and CMV proteases, changes are also seen in the A2 helix position with respect to A6. In HSV-2 protease [SEQ ID NO: 4], hydrogen bonds are formed between Ala 98 on A2 to both Ser 215 -OH (2.7 Å) and Asn 219 (3.3 Å) on A6. These two interactions and that between 211 and 207 [SEQ ID NO: 4] do not occur in either CMV or VZV proteases. In CMV and VZV proteases the dimer interface is along a two-fold crystallographic axis, which could cause subtle changes in the interactions between the monomers.

There are other notable structural differences between HSV-2 and HSV-1 and those of the CMV and VZV proteases discussed below. One difference is in the segment between A2 and A3. In the structure of CMV protease, this segment assumed a "closed" conformation, making intra-molecular contacts and being part of the dimer interface. In the VZV protease structure, this segment adapts a completely "open" conformation, interacting only with another symmetry-related molecule to form a different dimer interface. In the liganded and unliganded HSV-2 protease structures, part of this loop is disordered but is a close distance (9 Å) to another symmetry-related molecule. This suggests that HSV-2 and VZV proteases may be capable of using this segment to form higher order oligomers.

It has been reported that CMV protease dimerization is important for maintaining the activity of the protease [Darke, et al., *J. Biol. Chem.*, 271, pp. 7445-7449 (1996)]. In fact, the Darke publication and S. Margosiak, et al. *Biochemistry* 35, 5300-5307 (1996) have shown that the CMV protease molecules self-associate to form dimers in solution with a monomer-dimer equilibrium constant of approximately  $10^{-6}$  M and that the dimeric form of the protease is the only active species. There is now a similar report on the dimer formation of HSV-1 protease [Schmidt, U. and Darke, P. L., *J. Biol. Chem.*, 272, pp. 7732-7735 (1997)], but with a similar dimerization response in the presence of glycerol [Darke et al., *J. Biol. Chem.*, 270: 22697-22700 (1995)]. Furthermore, enzymatic assays are often reported in the presence of aggregation-promoting reagents such as glycerol or citrates [Burck et al., Hall et al., both cited above].

Based on the present structure it is difficult to determine why the dimer is required for activity, since the interface is not near either active site. The addition of the covalent inhibitor DIP does not appear to alter the dimer interface. However, rearrangements of the helices at the interface in the absence of a dimer could have profound effects on the conformation in the active site region. Thus, dimerization is believed to stabilize the conformation of helix A6.

## 2. The CMV and VZV Proteases

Unlike the structures of other serine proteases having two distinct b-barrel domains, the structures of each of the herpes proteases herein have a single domain.

The overall fold of the CMV and VZV monomers can be best described as a 7-stranded  $\beta$ -barrel core which in CMV is decorated with seven  $\alpha$ -helices on three sides (see Figs. 32A and 32B) and in VZV is decorated with eight  $\alpha$ -helices (Figs. 33A and 33B). The core  $\beta$ -barrel can be classified as an orthogonal packed  $\beta$ -barrel as described in detail by C. Chothia & J. Janin, *Biochemistry*, 21: 3955-3965 (1982). Of the seven helices of CMV, three were found between strand  $\beta$ 4 and  $\beta$ 5, and four helices after strand  $\beta$ 7 (Fig. 34).

Certain features of the CMV protease and VZV protease barrels are quite distinct. First, the CMV protease and VZV protease barrels contain two parallel strands (Figs. 34 and 35) and thus are mixed  $\beta$ -barrels, while most of the

orthogonal packed  $\beta$ -barrels are formed exclusively from anti-parallel strands. Second, strand  $\beta 3$  [aa67-78 of CMV (SEQ ID NO: 1) and aa57-67 of VZV (SEQ ID NO: 5)] are  $\beta$ -bends that close one corner of the barrel, but the other corner lacks this kind of classical closure and is maintained by only two hydrogen bonds between strand B5 [aa130-135 of CMV (SEQ ID NO: 1) and aa118-123 of VZV (SEQ ID NO: 5)] and B7 [aa169-175 of CMV (SEQ ID NO: 1) and aa151-157 of VZV (SEQ ID NO: 5)].

Interestingly, the N-terminal  $\beta$ -barrel of trypsin's prototype serine protease is an anti-parallel  $\beta$ -barrel that is also orthogonally packed. However, superposition of the CMV protease, VZV protease and trypsin barrels did not reveal any further resemblance, and show the enzyme active sites of the CMV and VZV proteases are at completely different regions of the fold than the active site of trypsin. Moreover, the CMV and VZV  $\beta$ -strands (Fig. 34 and Fig. 35) are arranged differently than in trypsin. For example, the first four strands of CMV and VZV proteases (B1, B2, B3 and B4) form a typical Greek Key motif, while those in trypsin do not. Therefore, it is reasonable to conclude that the CMV protease and VZV protease barrels are evolutionally unrelated to other, non-herpes serine proteases. The overall fold is also unique to CMV, VZV and other herpes proteases.

An intriguing dimer interface in CMV and VZV proteases has been identified around two-fold crystallographic axes (Figs. 36A and 36B, 37). The dimer interface is mainly made-up of a set of four helices (A1, A2, A3 and A6) of one monomer that surrounds helix A6 of the other monomer, where the two symmetry-related A6 helices are parallel (Fig. 37). The dimer interface is predominantly hydrophobic, involving many side chain van der Waals interactions for residues such as phenylalanines, leucines and valines. Despite the tight packing in the crystal, this dimer interface is much more significant than that of other inter-molecular interfaces within the crystal. The arrangements of the helices and the extent of the interface seem to suggest that this is not a simple coincidence of crystal packing.

The dimer interface is of importance in maintaining the activity of the proteases. The calculated interface area is between 850 (Connolly) to 1300  $\text{\AA}^2$  (GRASP) from the crystal structures of CMV and VZV proteases. As for HSV-2 and HSV-1, from the crystal structures, it is noted that the dimer interface is not in the immediate vicinity of the active site. Also, the active sites of the two monomers are quite distant from each other (Fig. 36B and Fig. 37).

Although the dimer interfaces in VZV and CMV proteases are similar, there are notable differences in their structures. Helices A6 from both monomers were almost parallel in the structure of CMV protease, but helix A6 is twisted about  $30^\circ$  in the VZV protease structure (Fig. 37). The helix A6 in VZV protease has one more turn at the N-terminal, and the loops connecting A5 and A6 are quite different in the two structures (Fig. 1). The biggest difference resides in the segment containing the small helix A2. In the structure of CMV protease, this segment assumed a "closed" conformation, making intra-molecular contacts and being part of the dimer interface. However, in the VZV protease structure, this segment adapts a completely "open" conformation, interacting only with another symmetry-related molecule to form a different dimer interface (Fig. 37). This suggests that VZV protease may be capable of using this segment to form higher order oligomers. By analogy with CMV protease, the VZV protease dimer is essential for enhanced catalytic activity. In the absence of a dimer, the rearrangements of helices involved in the interface may have profound effects on the conformation in the active site region. For example, the A6 helix may move toward the active site cavity and therefore affect the positioning of residues in the active site or it may simply block access to the substrate.

#### E. The novel active site

The catalytic mechanism of classical serine proteases involves an active site triad composed of a serine, histidine and an aspartic acid. However, although several prior art studies focused on mutagenesis of aspartic and glutamic acids of herpes proteases, none led to a correct identification of the third member of the herpes protease catalytic triad.

The crystal structure of liganded and unliganded HSV-2 protease reveals an active site composed of a serine (Ser 129), a histidine (His 61), and a third residue also a histidine (His 148) (Fig 38A,B), sequences which are conserved in all known herpes proteases (Fig. 1). The crystal structure of HSV-1 protease reveals an identical active site (Fig. 39A). The crystal structure is in agreement with early protease inhibition experiments on HSV-1 protease (sharing 90% sequence identity and identical numbering to HSV-2 protease) that identified HSV-1 as a serine protease and that substitution of His 148 and His 61 abolished [SEQ ID NO: 3] enzymatic activity (Liu & Roizman, Dilanni I). Similar studies on CMV protease had also demonstrated the homologous residues to His 61 and Ser 129 to be essential (Stevens, Welch I). Although several studies focused on mutagenesis of aspartic and glutamic acids of herpes proteases, none led to a correct identification of the third member of the catalytic triad. Figure 38A shows the DIP molecule covalently bound to Ser 129 [SEQ ID NO: 4]. This is consistent with mutagenesis and chemical modification studies that identified Ser 129 as the active site nucleophile in HSV-1 protease [SEQ ID NO: 3] (Dilanni II).

As with HSV-2 and HSV-1 proteases, the crystal structure of the CMV and VZV proteases of the invention reveals a novel active site containing a serine [Ser 132 for CMV (SEQ ID NO: 1) and Ser 120 for VZV (SEQ ID NO: 5)] and a histidine [His 63 for CMV (SEQ ID NO: 1) and His 52 for VZV (SEQ ID NO: 5)], with the third member of the catalytic triad being a histidine [His 157 for CMV (SEQ ID NO: 1) and His 139 for VZV (SEQ ID NO: 5)] instead of aspartic acid. Mutagenesis and chemical modification studies had identified Ser 132/Ser120 [SEQ ID NO: 1 and 5, respectively] and His 63/His52 [SEQ ID NO: 1 and 5, respectively] as part of the catalytic triad [Welch I, Stevens *et al.*, cited above]. Both

residues are absolutely conserved in all herpes proteases (Fig. 1).

### 1. HSV-2 and HSV-1 Proteases

5 The active site of DIP-liganded HSV-2 protease shows a network of hydrogen bonding between the enzyme active site, the ligand, and two central water molecules (Wat1 and Wat2 (Fig. 38A)). The crucial elements of the HSV-1 and HSV-2 active sites are strikingly similar to trypsin even though the two HSV enzymes share little sequence homology with trypsin and the overall tertiary structures are completely different to trypsin. An overlay of the catalytic triad of  $\gamma$ -chymotrypsin bound to mono-isopropyl phosphate (MIP) with that of the DIP-liganded HSV-2 structure (Fig. 38C), and an overlay of the catalytic triad of trypsin (BPD code 1SGT) with that of the HSV-1 protease structure shows this similarity (Fig. 39B). In Fig. 38C, significant overlap is seen between the peptide backbone stabilizing the P=O oxygen of DIP, the catalytic serine residues, and His 61 and His 57 side chains of  $\gamma$ -chymotrypsin and DIP-liganded HSV-2 protease, respectively. Figure 38C also reveals the overlap of Asp 102 and His 148 [SEQ ID NO: 4], supporting the role of this histidine in catalysis, despite its apparent lack of hydrogen bonds in this structure. Similar results are revealed in Fig. 39B. Not only does this confirm the role of His 148 [SEQ ID NO: 3 and 4] in catalysis, but it also suggests the possibility of converting the HSV-2 and HSV-1 enzymes into those having a normal catalytic triad by replacing His 148 [SEQ ID NO: 3 and 4] with an aspartic acid.

Because of the presence of the covalently bound inhibitor DIP, His 61 does not appear to hydrogen bond to Ser 129 [SEQ ID NO: 4] but instead maintains a close interaction with Ser 131 (2.5 Å). This can be compared to a slightly different hydrogen bonding network in apo CMV protease (Fig. 38D) where His 61 has hydrogen bonds to both Ser 129 and His 148 and Ser 131 clearly interacts with His 148 [SEQ ID NO: 1]. This is similar to the hydrogen bonding network in the  $\gamma$ -chymotrypsin/MIP structure where Ser 214 maintains a close hydrogen bond to Asp 102 (Fig. 38C) [SEQ ID NO: 3]. Despite its location in the active site, Ser 131 has been found to be nonessential for catalysis in CMV protease [SEQ ID NO: 1] (Welch I). This position is also an Ala residue in HSV-1 (Fig 1). The average B factor of the side chain atoms of His 61 in the DIP-liganded HSV-2 protease [SEQ ID NO: 4] is 51 Å<sup>2</sup>, more than twice the average B-factor of the structure, indicating that it is mobile. A rotation about the C $\beta$ -C $\gamma$  bond could allow a hydrogen bond between these two residues. These two rotations would present a hydrogen bond of 3.0 Å between His 148 and His 61, about the same distance as in the uncomplexed CMV structure, and thus an alternative set of hydrogen bonds.

30 The active site of HSV-1 protease is very similar to that of HSV-2 protease with some minor differences (Fig. 39C). These differences are most likely because the HSV-2 protease had a covalently bound DIP inhibitor (not shown in the figure, for clarity) bound to the Ser 129 which prevented a hydrogen bond between Ser 129 and His 61 [SEQ ID NO: 4]. This hydrogen bond is present in the HSV-1 protease structure in which the imidazole ring has turned by about 90° to accommodate this hydrogen bond. Also, position 131 is an Alanine in HSV-1 protease [SEQ ID NO: 3] and thus cannot maintain any hydrogen bonds to either His 61 or His 148 as has been seen in HSV-2 [SEQ ID NO: 4] and the other herpes proteases. There is a slightly different hydrogen bonding network in the unliganded (or apo) CMV protease structure (Fig. 39D) where the equivalent residue of His 61 has hydrogen bonds to both Ser 129 and His 148 and Ser 131 clearly interacts with His 148 [SEQ ID NO: 1].

Another absolutely conserved residue in both HSV-1 and HSV-2 is Cys 152 [SEQ ID NO: 4] which is within the vicinity of the active site. It is also conserved and in an identical position in trypsin. However, it has limited contact with the DIP ligand (Fig. 38A): the C152 Cy atom maintains a van der Waals contact (3.8 Å) to a Ser 129 methyl group. Thus, it is difficult to imagine it being a suitable proton acceptor because of its nature and position. Also, this cysteine is not essential for catalytic activity in HSV-1 protease (Liu & Roizman III), or CMV protease (Welch I).

45 An oxyanion hole for DIP-liganded HSV-2 protease can be identified in the present invention. Such an oxyanion hole for HSV-1 protease can also be identified based on its nearly identical structure to the HSV-2 proteases (Fig. 39C). An oxyanion hole is that portion of the protease which provides an environment for the stabilization of the tetrahedral intermediate. In DIP-liganded HSV-2 protease, the amide nitrogen of Arg 156 and Wat1 stabilize the P=O oxygen of the DIP and define the oxyanion hole of the enzyme (Fig. 38A). Wat1 is stabilized by hydrogen bonds to Wat2 (2.7 Å) and Val128 (2.8 Å). Correspondingly, Wat 2 is held by hydrogen bonds to backbone atoms of Leu 130 (2.9 Å) and Leu127 (2.9 Å) and Arg 157N $\epsilon$  (3.2 Å). The alignment with CMV protease (Fig. 38D) also shows a single water molecule in the active site region of this enzyme, closely overlapping with Wat2 in the HSV-2 protease structure. This water molecule maintains the same protein backbone hydrogen bonds as does Wat2 in HSV-2 and could help hold the side chain of Arg 157 in place. In HSV-1 protease, the backbone atoms at Arg156 are the same as in DIP-liganded HSV-2 protease, making it likely this residue also helps define the oxyanion hole in HSV-1 protease. Arg156 and Arg 157 are absolutely conserved in all herpes proteases and present an overall positive charge near the oxyanion hole. The stability of this region is reflected here where Arg 157 makes two hydrogen bonds to backbone atoms of Leu 130 and Leu 38 both absolutely conserved in all herpes proteases (Fig. 1). The alignment with  $\gamma$ -chymotrypsin shows how close the P=O oxygen of MIP is stabilized by a hydrogen bond to the amide nitrogen of Gly 193. In DIP-liganded HSV-2 protease, the amide nitrogen of Arg 156 closely corresponds to that of Gly 193 even though the overall structure of the two enzymes

is completely different (Fig. 38C).

The active site of liganded and unliganded human HSV-2 protease and HSV-1 protease sit at a very shallow and mostly exposed region of the protease (Figs. 27A,B,C; 29A,B,C; Fig. 28A, 28B, 30A, 30B). Shallowness of the active site cavity is not really surprising given that the scissile bond (the bond which gets cleaved) recognized by all herpes proteases is between two small amino acid residues (Ala-Ser). Missing around the active site cavity are amino acid residues 134-140 [SEQ ID NO: 3 and 4], that are part of a surface loop. Interestingly, a mutant with a five residue deletion in the corresponding loop in CMV protease residue was shown to be fully active, but with altered substrate specificity [Welch I, cited above]. Given this loop's proximity to the active site cavity, it may be a flexible flap that is involved in substrate recognition.

Since the CMV protease structure misses two large loops near the active site, it was difficult to speculate about the substrate binding mode of the enzyme. With the liganded and unliganded HSV-2 protease and HSV-1 protease structures, the missing loop containing residues 32-54 [SEQ ID NO: 3 and 4] becomes ordered, possibly having a role in substrate recognition (Fig. 27A, Fig. 28A). There are two grooves, or depressions near the active site. One of the grooves is deep and wide, found in a region that is reminiscent of the S' subsites of classical serine proteases. One side of the groove is delineated by the active site residues while the other is formed by a side of helix A6, a critical structural feature of the enzyme. The other groove is relatively narrow and is formed by B5, including the catalytic triad, on one side and the other made by the small loop of 154-160 [SEQ ID NO: 3 and 4] which includes the conserved GRR sequence (Fig. 1). This region is also in a position that is not very different from the unprimed (S) subsites in classical serine proteases. The substrate peptide (at least P2-P4) could be inserted into the groove with its main chain forming an antiparallel  $\beta$ -sheet with strand B5 and B6. Of course, structural studies of enzyme-substrate analog complexes are needed for proving this model.

## 2. CMV Protease and VZV Protease Active Sites

None of the aspartic or glutamic acids is absolutely conserved in all herpes proteases. Glu 122 was proposed as a member of the catalytic triad [Cox *et al.*, cited above] for CMV protease. However, it is found to be distant from the catalytic site in the CMV protease crystal structure. This glutamic acid is buried near the C-terminus of the protein, making a salt bridge with Lys 255 of CMV protease (Glu 122 OE1-Lys 255 NZ, 2.7 Å) and a hydrogen bond with the backbone nitrogen of Asp 118 of CMV protease (Glu122 OE2-Asp 118 N, 3.1 Å) [SEQ ID NO: 1]. Therefore, its importance to the protease can only be attributed to its role in maintaining the overall structure of the protease, rather than being directly involved in the catalytic machinery.

Although His 157 (for CMV, SEQ ID NO: 1) and His 139 (for VZV, SEQ ID NO: 5) are absolutely conserved among all herpes proteases, and mutagenesis of this histidine was shown to abolish enzymatic activity in HSV-1 [Liu II, cited above] and CMV proteases in HSV-1 [Welch I, cited above], no one has suggested that it has a role as the third member of the catalytic triad. Abolition of enzymatic activity does not necessarily necessitate involvement in catalytic activity but could be a result of changes in protein conformation.

As expected, the O $\gamma$  atom of Ser 132 for CMV protease [SEQ ID NO: 1] (Ser 120 for VZV protease [SEQ ID NO: 5]) is found to be in the vicinity of His 63 for CMV protease (His52 for VZV protease), with a distance of 3.3 Å from its N $\epsilon$ 2 nitrogen for CMV protease (3.6 Å for VZV protease). Given the presence of about 0.4 Å coordinates error and the absence of a substrate in the active site, the 3.3 Å distance for CMV (3.6 Å for VZV) does not preclude these two residues from being the catalytic residues. Surprisingly, the conserved second histidine (His 157 for CMV protease [SEQ ID NO: 1] and His 139 for VZV protease [SEQ ID NO: 5]) is hydrogen bonded to the side chain of His 63 (His 63 N $\delta$ 1-His 157 N $\epsilon$ 2, 3.2 Å) for CMV protease and His 52 (His 52 N $\delta$ 1-His 139 N $\epsilon$ 2, 3.2 Å) for VZV protease, making it the third member of the catalytic triad. In CMV protease [SEQ ID NO: 1], the only acidic residue in the vicinity is Asp 65, with its O $\delta$ 1 atom 3.9 Å away from the N $\delta$ 1 nitrogen of His 157. In VZV protease [SEQ ID NO: 5], a basic residue Lys 54 replaces the Asp 65 of CMV protease, with its O $\delta$ 1 atom 5.1 Å away from the N $\epsilon$ 2 nitrogen of His 139, too far to influence catalysis in VZV protease.

In the CMV protease crystal structure, Asp 65 forms a salt-bridge with Arg 109 of a neighboring, symmetry-related molecule other than the aforementioned dimer (Asp 65 O $\delta$ 2-Arg 109 NH1, 2.8 Å). Therefore, in the absence of this salt-bridge in solution, Asp 65 side chain could readily move to hydrogen-bond with His 157 and act as a proton acceptor in what can be described as a catalytic tetrad. In the active site, Asn 60 is also found to interact with His 157 (Asn 60 N $\delta$ 2-His 157 N $\delta$ 1, 3.4 Å; Asn 60 N $\delta$ 2-His 157 N $\epsilon$ 2, 3.7 Å), but it is difficult to imagine it being a suitable proton acceptor because of its nature and position. Therefore, the active site of CMV protease [SEQ ID NO: 1] either consists of a novel triad of Ser 132, His 63 and His 157, or a unique tetrad consisting of Ser 132, His 63, His 157 and Asp 65 that has also never been reported previously (Fig. 41A). In the "catalytic tetrad", His 157 acts as an extra component in this novel "relay" proton transfer mechanism. However, the lack of sequence conservation for Asp 65 (Fig. 1) indicates that this tetrad is not a general model for herpes proteases.

In the active site of the VZV protease [SEQ ID NO: 5] crystal structure, Cys 143 is also found to interact with Ser 120 (Cys 143 S $\gamma$ -Ser 120 O $\gamma$  4.8 Å; Ser 122 O $\gamma$ -His 139 C $\epsilon$ 1, 3.0 Å; Cys 143 S $\gamma$ -His 52 N $\epsilon$ 2 6.0 Å; and Ser 122 O $\gamma$ -

His 139 N $\delta$ 1 2.9 Å), but it is difficult to imagine it being a suitable proton acceptor because of its nature and position. Therefore, the active site of VZV protease consists of a novel triad of Ser 120, His 52 and His 139 of SEQ ID NO: 1 (Fig. 40A).

Overlay of Ser 132 and His 63 of CMV protease (and Ser120 and His52 of VZV protease) on Ser 195 and His 57 of the classical serine protease triad in trypsin reveals that His 157 of CMV protease [SEQ ID NO: 1] and His139 of VZV protease [SEQ ID NO: 5] can be superimposed almost perfectly onto Asp 102 of trypsin (Fig. 41B and Fig. 40B). Not only does this confirm the role of CMV protease His 157 and VZV protease His 139 in catalysis, but it also suggests the possibility of converting this enzyme into one having a normal catalytic triad by replacing His 157 in CMV protease or His 139 in VZV protease with an aspartic acid, which may also require substituting Asp 65 by a non-acidic residue.

In the same overlay, despite a totally different tertiary structure, many interesting conservations can be identified that seem to present a case of convergent evolution. First, Cys 161 for CMV protease and Cys 143 for VZV protease are at an identical position to Cys 42 of trypsin (Fig. 41B and Fig. 40B), making van der Waals interactions with the catalytic residues (Cys 161 S $\gamma$ -Ser 132 O $\gamma$ , 3.6 Å; Cys 161 S $\gamma$ -His 63 N $\epsilon$ 2, 4.4 Å in CMV). This surprising conservation, as well as the fact that Cys 161 in CMV protease [SEQ ID NO: 1] and Cys 143 in VZV protease [SEQ ID NO: 5] are absolutely conserved in all herpes proteases, seems to suggest an important role for this amino acid, although mutagenesis studies had shown that it is not essential for the protease activity [Welch I, cited above]. Another similar scenario is found at Ser 134 in CMV protease and Ser 122 in VZV protease, which appear to be at identical positions to Ser 214 of trypsin (Fig. 41B and Fig. 40B).

In the CMV structure of the present invention, Ser 134 interacts strongly with His 157 (Ser 134 O $\gamma$ -His 157 N $\epsilon$ 2, 2.6 Å; Ser 134 O $\gamma$ -His 157 N $\delta$ 1, 3.0 Å) by forming a hydrogen bond. In the VZV structure of the invention, Ser 122 interacts strongly with His 139 (Ser 122 O $\gamma$ -His 139 N $\epsilon$ 2, 2.5 Å) by forming a hydrogen bond. In both CMV and VZV structures, Ser 214 also interacts strongly with Asp 102 in trypsin by forming a hydrogen bond. However, the importance of Ser 134 in CMV protease to the catalytic activity has been undermined by mutagenesis studies [Welch I, cited above], and also the fact that it is an alanine in other herpes virus proteases.

A possible oxyanion hole for the CMV and VZV proteases also exists. In trypsin, the oxyanion is held by the backbone nitrogen atoms of Gly 193 and Ser 195. In the similar region of CMV protease, the construction of the oxyanion hole cannot be fully imitated by the G-X-S-G-G [SEQ ID NO: 14] motif because the backbone arrangements are completely different. However, the main chain nitrogen atom of Arg 165 in CMV protease and Arg 147 in VZV protease is at a nearly identical position as Gly 193 N in trypsin (Figs. 41A and 41B and Figs. 40A and 40B). Also found in the vicinity is a water molecule held by the side chain of Arg 166 and interacting with Leu 20 and Leu 133 (H<sub>2</sub>O-Arg 166 NH1, 2.7 Å; H<sub>2</sub>O-Arg 166, NZ, 3.2 Å; H<sub>2</sub>O-Leu 20 O, 2.6 Å; H<sub>2</sub>O-Leu 133 N 3.0 Å). The oxyanion in the structure defined herein may be held only by Arg 165 N, by Arg 165 N and the H<sub>2</sub>O molecule. Considering the fact that the two arginines (165 and 166) are absolutely conserved among all herpes proteases (Fig. 1), this general region is suitable for being the oxyanion pocket of CMV [SEQ ID NO: 1] and VZV [SEQ ID NO: 5] proteases.

The active sites of human CMV and VZV proteases sit at a very shallow and mostly exposed region of the protease (Figs. 36B, 41A, 37, 40A). Shallowness of the active site cavity is not really surprising given that the scissile bond (the bond which gets cleaved) recognized by all herpes proteases is between two small amino acid residues (Ala-Ser). Missing around the active site cavity are amino acid residues 143-153 in CMV protease [SEQ ID NO: 1] and aa 139-154 in VZV protease [SEQ ID NO: 5], that are part of a surface loop. This loop contains the so called inactivation or internal (I) site, a cleavage site between Ala 143 and Ala 144 of native human CMV protease as described in Welch I, cited above.

Residue 143 [SEQ ID NO: 1] of the CMV protease of this invention has been mutated to valine to eliminate such processing. Also, it is not clear whether cleavage at the I site is a result of auto-processing or not. Interestingly, a mutant with a five residue deletion around residue 143 was shown to be fully active, but with altered substrate specificity [Welch I, cited above]. Given this loop's proximity to the active site cavity, it may be a flexible flap that is involved in substrate recognition and probably is ordered upon binding of ligands. Similarly, the missing loop containing residues 25-55 may also become ordered upon ligand binding. This is supported by the fact that a mutation of Glu22 in simian CMV protease (corresponding to Glu31 in the human enzyme), has shown altered substrate specificity [Welch I, cited above].

Since the CMV protease structure misses two large loops near the active site, it was difficult to speculate about the substrate binding mode of the enzyme. With the VZV protease structure, the missing loop containing residues 23-45 of SEQ ID NO: 5 becomes ordered, and the structure clearly defines two grooves near the active site that could be important for substrate recognition (Fig. 37). One of the grooves is deeper and wider, and is found in a region that is reminiscent of the S' subsites of classical serine proteases. One side of the groove is delineated by the active site residues, while the other is formed by a side of helix A6, which is also a critical structural feature of the enzyme and will be discussed later. The other groove is relatively narrow. The  $\beta$ -strand B5, including the catalytic triad, is on one side of the shallow depression. The other side is formed by the conserved GRR sequence (Fig. 1) as well as the loop immediately prior to helix AA. This region is also in a position that is not very different from the unprimed (S) subsites in classical serine proteases. Strand B5 being almost parallel to this groove suggests that the substrate peptide (at least P2-P4) could be inserted into the groove with its main chain forming an antiparallel  $\beta$ -sheet with strand B5 and B6. Moreover, several rather exposed hydrophobic residues in the AA loop could also make important interactions with the substrate

protein. Of course, structural studies of enzyme-substrate analog complexes are needed for proving this model.

Given the conservation of amino acid sequence and substrate specificity between CMV, VZV, HSV-1, HSV-2 and other herpes proteases, the structures described herein represent that of the entire family of herpes proteases. These structures are clearly useful in the structure-based design of protease inhibitors, which may be used as therapeutic agents against viral disease. The discovery of the herpes protease catalytic triad, and the catalytic tetrad, permits the design of potent, highly selective protease inhibitors.

#### F. Mutants and Derivatives

The invention further provides homologues, co-complexes, mutants, derivatives and fragments of the herpes protease crystal structure of the invention.

The term "homologue" means a protein having at least 25% amino acid sequence identity with herpes protease or any functional domain of herpes protease. See. Fig. 1.

The term "co-complex" means herpes protease or a mutant or homologue of herpes protease in covalent or non-covalent association with a chemical entity or compound.

The term "mutant" refers to a herpes protease polypeptide, i.e., a polypeptide displaying the biological activity of wild-type protease activity, characterized by the replacement of at least one, or more, amino acids from the wild-type protease sequence. Such a mutant may be prepared, for example, by expression of herpes protease cDNA previously altered in its coding sequence by oligonucleotide-directed mutagenesis.

Herpes protease mutants may also be generated by site-specific incorporation of unnatural amino acids into herpes protease proteins using the general biosynthetic method of C. J. Noren et al, *Science*, 244:182-188 (1989).

In this method, the codon encoding the amino acid of interest in wild-type herpes protease is replaced by a "blank" nonsense codon, TAG, using oligonucleotide-directed mutagenesis. A suppressor tRNA directed against this codon is then chemically aminoacylated *in vitro* with the desired unnatural amino acid. The aminoacylated tRNA is then added to an *in vitro* translation system to yield a mutant herpes protease enzyme with the site-specific incorporated unnatural amino acid.

Selenocysteine or selenomethionine may be incorporated into wild-type or mutant herpes protease by expression of herpes protease-encoding cDNAs in auxotrophic *E. coli* strains [W. A. Hendrickson et al, *EMBO J.*, 9(5):1665-1672 (1990)]. In this method, the wild-type or mutagenized herpes protease cDNA may be expressed in a host organism on a growth medium depleted of either natural cysteine or methionine (or both) but enriched in selenocysteine or selenomethionine (or both).

The term "heavy atom derivative" refers to derivatives of herpes protease produced by chemically modifying a crystal of herpes protease. In practice, a crystal is soaked in a solution containing heavy metal atom salts, or organometallic compounds, e.g., lead chloride, gold thiomalate, thiomersal or uranyl acetate, which can diffuse through the crystal and bind to the surface of the protein. The location(s) of the bound heavy metal atom(s) can be determined by X-ray diffraction analysis of the soaked crystal. This information, in turn, is used to generate the phase information used to construct three-dimensional structure of the enzyme [T. L. Blundell and N. L. Johnson, *Protein Crystallography*, Academic Press (1976). See. Example I.

The term "fragment", particularly as used in connection with protease fragments, refers to a protease of the invention which contains at least the catalytic active site of the protease, but less than the full length protease. Desirably, the fragment is characterized by a catalytic active site which has the same crystal structure as the active site in the full-length protease. However, a fragment of the invention is not so limited. Such a fragment may contain N-terminal, C-terminal or internal deletions of the protease. Particularly desirable are fragments which are N-terminally truncated proteases. It is currently anticipated that such fragments provide superior resolution or are more easily crystallized.

## II. Methods of Identifying Inhibitors of the Novel Protease Crystalline Structure

Another aspect of this invention involves a method for identifying inhibitors of a herpes protease characterized by the crystal structure and novel active site described herein, and the inhibitors themselves. The novel protease crystal structure of the invention permits the identification of inhibitors of protease activity. Such inhibitors may bind to all or a portion of the active site of the herpes protease; or even be competitive non-competitive, or uncompetitive inhibitors; or interfere with dimerization by binding at the interface between the two monomers. Once identified and screened for biological activity, these inhibitors may be used therapeutically or prophylactically to block protease activity, and thus, herpes viral replication latency, reactivation and/or infection.

One design approach is to probe the herpes protease crystal of the invention with molecules composed of a variety of different chemical entities to determine optimal sites for interaction between candidate herpes protease inhibitors and the enzyme. For example, high resolution X-ray diffraction data collected from crystals soaked in or co-crystallized with other molecules allows the determination of where each type of solvent molecule sticks. Molecules that bind tightly to those sites can then be further modified and synthesized and tested for their herpes protease inhibitor activity [J. Travis,

Science, 262:1374 (1993)].

This invention also enables the development of compounds that can isomerize to short-lived reaction intermediates in the chemical reaction of a substrate or other compound that binds to or with herpes protease. The time-dependent analysis of structural changes in herpes protease during its interaction with other molecules is permitted. The reaction intermediates of herpes protease can also be deduced from the reaction product in co-complex with herpes protease. Such information is useful to design improved analogues of known herpes protease inhibitors or to design novel classes of inhibitors based on the reaction intermediates of the herpes protease enzyme and herpes protease inhibitor co-complex. This provides a novel route for designing herpes protease inhibitors with both high specificity and stability.

Another approach made possible by this invention, is to screen computationally small molecule data bases for chemical entities or compounds that can bind in whole, or in part, to the herpes protease enzyme. In this screening, the quality of fit of such entities or compounds to the binding site may be judged either by shape complementarity or by estimated interaction energy [E. C. Meng et al, *J. Comp. Chem.*, 13:505-524 (1992)].

Because herpes protease may crystallize in more than one crystal form, the structure coordinates of herpes protease, or portions thereof, as provided by this invention are particularly useful to solve the structure of those other crystal forms of herpes protease. They may also be used to solve the structure of herpes protease mutants, herpes protease co-complexes, or of the crystalline form of any other protein with significant amino acid sequence homology to any functional domain of herpes protease.

One method that may be employed for this purpose is molecular replacement. In this method, the unknown crystal structure, whether it is another crystal form of herpes protease, a herpes protease mutant, or a herpes protease co-complex, or the crystal of some other protein with significant amino acid sequence homology to any functional domain of herpes protease, may be determined using the herpes protease structure coordinates of this invention as provided in Figs. 1-26. This method will provide an accurate structural form for the unknown crystal more quickly and efficiently than attempting to determine such information *ab initio*.

Thus, the protease structure provided herein permits the screening of known molecules and/or the designing of new molecules which bind to the protease structure, particularly at the active site, via the use of computerized evaluation systems. For example, computer modelling systems are available in which the sequence of the protease, and/or the protease structure (i.e., atomic coordinates of CMV, VZV, HSV-2, or HSV-1 proteases and/or the atomic coordinates of the active site cavity, bond angles, dihedral angles, distances between atoms in the active site region, etc. as provided by Figs. 1-26), may be input. Alternatively, the catalytic site domain crystal structure of a protease of the invention or another fragment of the protease may be input into computer readable form. Thus, for DIP-liganded HSV-2 protease, a machine readable medium may be encoded with data representing the coordinates of Figs. 2, 3, 8, 9, 11, and 12; or Figs. 2, 3, 8, 9, 14, and 15 (Figs. 4 and 5 may be substituted for Figs. 2 and 3 in the process). Similarly, for HSV-1 protease, a machine readable medium may be encoded with data representing the coordinates of Figs. 6, 10 and 13; or Figs. 6, 10 and 16 (as noted Fig. 7 may be substituted for Fig. 6 in this process). For CMV protease, a machine readable medium may be encoded with data representing the coordinates of Figs. 17, 18 and 19; or Figs. 17, 18 and 20 (Fig. 21 may be substituted for Fig. 17 in this process). For VZV protease, a machine readable medium may be encoded with data representing the coordinates of Figs. 22, 24 and 25; or Figs. 22, 24 and 26 (Fig. 23 may be substituted for Fig. 22 in this process). The computer then generates structural details of the site into which a test compound should bind thereby enabling the determination of the complementary structural details of said test compound.

More particularly, the design of compounds that bind to or inhibit herpes protease according to this invention generally involves consideration of two factors. First, the compound must be capable of physically and structurally associating with the herpes protease and, particularly, with the active site thereof. Non-covalent molecular interactions important in the association of herpes protease with its ligands include hydrogen bonding, van der Waals and hydrophobic interactions.

Second, the compound must be able to assume a conformation that allows it to associate with herpes protease. Although certain portions of the compound will not directly participate in this association with herpes protease, those portions may still influence the overall conformation of the molecule. This, in turn, may have a significant impact on potency. Such conformational requirements include the overall three-dimensional structure and orientation of the chemical entity or compound in relation to all or a portion of the binding site, e.g., active site or accessory binding site of herpes protease, or the spacing between functional groups of a compound comprising several chemical entities that directly interact with herpes protease.

The potential inhibitory or binding effect of a chemical compound on herpes protease may be analyzed prior to its actual synthesis and testing by the use of computer modelling techniques. If the theoretical structure of the given compound suggests insufficient interaction and association between it and herpes protease, synthesis and testing of the compound is obviated. However, if computer modelling indicates a strong interaction, the molecule may then be synthesized and tested for its ability to bind to herpes protease and inhibit using a suitable assay. In this manner, synthesis of inoperative compounds may be avoided.

An inhibitory or other binding compound of herpes protease may be computationally evaluated and designed by means of a series of steps in which chemical entities or fragments are screened and selected for their ability to associ-

ate with the individual binding pockets or other areas of herpes protease.

One skilled in the art may use one of several methods to screen chemical entities or fragments for their ability to associate with herpes protease and more particularly with the individual binding pockets of the herpes protease active site or accessory binding site. This process may begin by visual inspection of, for example, the active site on the computer screen based on the herpes protease coordinates in Figs. 2, 3, 6, 8-20, 22, and 24-26. Selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within a binding pocket of herpes protease. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics forcefields, such as CHARMM and AMBER.

Specialized computer programs may also assist in the process of selecting fragments or chemical entities. These include the GRID program available from Oxford University, Oxford, UK. [P. J. Goodford, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", *J. Med. Chem.*, 28:849-857 (1985)]; the MCSS program available from Molecular Simulations, Burlington, MA [A. Miranker and M. Karplus, "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method", *Proteins: Structure, Function and Genetics*, 11:29-34 (1991)]; the AUTODOCK program available from Scripps Research Institute, La Jolla, CA [D. S. Goodsell and A. J. Olsen, "Automated Docking of Substrates to Proteins by Simulated Annealing", *Proteins: Structure, Function and Genetics*, 8:195-202 (1990)]; and the DOCK program available from University of California, San Francisco, CA [I. D. Kuntz et al, "A Geometric Approach to Macromolecule-Ligand Interactions", *J. Mol. Biol.*, 161:269-288 (1982)]. Additional commercially available computer databases for small molecular compounds include Cambridge Structural Database, Fine Chemical Database, and CONCORD database [for a review see Rusinko, A., *Chem. Des. Auto. News*, 8:44-47 (1993)].

Once suitable chemical entities or fragments have been selected, they can be assembled into a single compound or inhibitor. Assembly may proceed by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of herpes protease. This would be followed by manual model building using software such as Quanta or Sybyl.

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include the CAVEAT program [P. A. Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in *Molecular Recognition in Chemical and Biological Problems*, Special Pub., Royal Chem. Soc. 78, pp. 182-196 (1989)], which is available from the University of California, Berkeley, CA; 3D Database systems such as MACCS-3D database (MDL Information Systems, San Leandro, CA) [see, e.g., Y. C. Martin, "3D Database Searching in Drug Design", *J. Med. Chem.*, 35:2145-2154 (1992)]; and the HOOK program, available from Molecular Simulations, Burlington, MA.

Instead of proceeding to build a herpes protease inhibitor in a step-wise fashion one fragment or chemical entity at a time as described above, inhibitory or other herpes protease binding compounds may be designed as a whole or "de novo" using either an empty active site or optionally including some portion(s) of a known ligand(s). Suitable methods describing such methods include the LUDI program [H.-J. Bohm, "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", *J. Comp. Aid. Molec. Design*, 6:61-78 (1992)], available from Biosym Technologies, San Diego, CA; the LEGEND program [Y. Nishibata and A. Itai, *Tetrahedron*, 47:8985 (1991)], available from Molecular Simulations, Burlington, MA; and the LeapFrog program, available from Tripos Associates, St. Louis, MO.

Other molecular modelling techniques may also be employed in accordance with this invention. See, e.g., N. C. Cohen et al, "Molecular Modeling Software and Methods for Medicinal Chemistry", *J. Med. Chem.*, 33:883-894 (1990). See also, M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", *Current Opinions in Structural Biology*, 2:202-210 (1992). For example, where the structures of test compounds are known, a model of the test compound may be superimposed over the model of the structure of the invention. Numerous methods and techniques are known in the art for performing this step, any of which may be used. See, e.g., P.S. Farmer, *Drug Design*, Ariens, E.J., ed., Vol. 10, pp 119-143 (Academic Press, New York, 1980); U.S. Patent No. 5,331,573; U.S. Patent No. 5,500,807; C. Verlinde, *Structure*, 2:577-587 (1994); and I. D. Kuntz, *Science*, 257:1078-1082 (1992). The model building techniques and computer evaluation systems described herein are not a limitation on the present invention.

Thus, using these computer evaluation systems, a large number of compounds may be quickly and easily examined and expensive and lengthy biochemical testing avoided. Moreover, the need for actual synthesis of many compounds is effectively eliminated.

Once identified by the modelling techniques, the protease inhibitor may be tested for bioactivity using standard techniques. For example, structure of the invention may be used in binding assays using conventional formats to screen inhibitors. Suitable assays for use herein include, but are not limited to, the enzyme-linked immunosorbent assay (ELISA), or a fluorescence quench assay. See, for example, the HSV-1, HSV-2, CMV, and VZV protease activity assays below. Other assay formats may be used; these assay formats are not a limitation on the present invention.

In another aspect, the protease structure of the invention permits the design and identification of synthetic compounds and/or other molecules which have a shape complementary to the conformation of the protease active site of the invention. Using known computer systems, the coordinates of the protease structure of the invention may be provided in machine readable form, the test compounds designed and/or screened and their conformations superimposed

on the structure of the protease of the invention. Subsequently, suitable candidates identified as above may be screened for the desired protease inhibitory bioactivity, stability, and the like.

Once identified and screened for biological activity, these inhibitors may be used therapeutically or prophylactically to block protease activity, and thus, herpes viral replication.

As used herein the term "natural product molecule" includes all non-synthetic products of nature and includes, but is not limited to, derivatives, extracts or homologs thereof, having, or containing, a bioactive component.

The following examples illustrate various aspects of this invention. These examples do not limit the scope of this invention which is defined by the appended claims.

#### 10 Example 1: Analysis of the Structure of the HSV-2 Protease

The HSV-2 protease (see Fig. 1, SEQ ID NO: 4) was cloned, expressed and purified as follows:

##### A. Expression, Purification and Crystallization

15 HSV-2 protease was expressed in *E. coli* including a 19-residue addition beyond its C-terminal alanine residue [SEKFKIWGAESAPHHHHH (SEQ ID NO: 15)]. The hexa-His tag (the six H) allows the high-quality purification of the protein using a  $\text{Ni}^{2+}$ -NTA chromatographic column. The construct also allows the protease to self-process by cleaving the peptide bond between the C-terminal alanine and the first added residue (Ser), thus producing a protein that  
20 has the same length as the authentic protease. The protease was further purified using Superdex 75 size exclusion and, if necessary, Q-Sepharose anion exchange chromatography. For the DIP-liganded HSV-2 protease, diisopropylfluorophosphate inhibitor (DFP) was added to the enzyme and incubated until >98% modification. Excess inhibitor was removed by Sephadex G-25 chromatography.  
The DIP-liganded HSV-2 protease was crystallized in 0.1 M NaAcetate buffer pH 5.0 and 10% PEG 4000 (50% w/v).  
25 Large crystals are approximately 0.7mm x 0.3mm x 0.2mm in size. The unliganded HSV-2 protease was crystallized in 0.1 M phosphate/Citrate buffer at pH 4.5, 20% PEG 8000. The crystals were 0.3 mm x 0.2 mm X 0.2 mm in size.

##### B. X-ray Diffraction Characterization

30 For the liganded and unliganded HSV-2 proteases a crystal was mounted in a sealed glass capillary with a small amount of mother liquor in each end of the capillary. The  $\text{CuK}_\alpha$  x-ray, having a wavelength of 1.54 Å, was generated by a Siemens-RU200 rotating anode machine operating at 50 KV x 95 mA electric power. The crystal was exposed to the  $\text{CuK}_\alpha$  x-ray, and the diffracted X-ray was collected by a Siemens multiwire area detector. The DIP-liganded HSV-2 protease crystal diffracted to 2.5 Å resolution. By registering the position and intensity of many tens of thousands of diffraction spots using the computer program XDS, [Kabsch, W., *J. Appl. Cryst.*, 21, pp. 916-924 (1988)] the crystal has been  
35 determined to be the orthorhombic space group  $P2_12_12$ , with  $a = 71.7$  Å,  $b = 87.4$  Å and  $c = 77.3$  Å. By established methods, an asymmetric unit was calculated to have two protein molecules. The crystal contains an estimated 45% solvent. The native data is 91% complete to 2.5 Å with an  $R_{\text{sym}} (\sum |I_i - \langle I \rangle| / \sum I_i)$  of 0.095. The unliganded HSV-2 protease crystal diffracted to 2.8 Å resolution with cell dimensions and space group identical to the DIP-liganded HSV-2 protease  
40 crystal. The native data is 94% complete to 2.8 Å with an  $R_{\text{sym}}$  of 0.095.

##### C. Heavy Atom Derivative:

45 Multiple isomorphous replacement (MIR) methods were used as one of the methods in order to obtain phase information of the diffraction data and to solve the three-dimensional atomic structure of the DIP-liganded HSV-2 protease. This involves the identification of derivative crystals containing specifically-bound heavy metal atoms. By testing various heavy metal compounds, the useful derivatives were prepared by soaking the native crystals with 0.1mM  $\text{KAuCN}$ , 0.4mM  $\text{LuCl}_3$ , 0.6mM  $\text{PrCl}_3$ , 0.2mM  $\text{YbS}_4$ , 0.5mM  $\text{GdCl}_3$ , and 0.2 mM  $\text{SmCl}_3$  for one to two days. The X-ray diffraction data of the derivative were then collected by the same methods described above. Data collection statistics for native  
50 and heavy atom derivatives are shown in Figure 42. Heavy atom positions were identified by difference Patterson and difference Fourier methods using the programs in the XtalView software package [McRee, D.E. *Practical Protein Crystallography*, (San Diego, Academic Press 1993). Heavy atom refinement and determination of an initial set of phases were carried out using the programs in the CCP4 suite [Collaborative Computational Project, Number 4, The CCP4 Suite: Programs for Protein Crystallography *Acta Crystallogr.* D50, 760-763 (1994)]. The program MLPHARE  
55 [Otwinowski, Z. *Isomorphous Replacement and Anomalous Scattering*, 80-86, Daresbury Laboratory, Warrington (1991)] was used for heavy atom phasing. Using the initial phases obtained through the MIR methods, a map of electron density within the crystal unit cell could be calculated. Because electrons are heavily distributed in the immediate vicinity of the centers of atoms, the positions of protein atoms are registered according to the electron density map. The resulting electron density map was interpretable but the phase information from MIR was improved with more phase

information derived from molecular replacement.

#### D. Molecular Replacement

5 A molecular replacement solution of the DIP-liganded HSV-2 protease was also identified with the program XPLOR [Brünger, A. T. *X-PLOR Version 3.1 A System for X-ray Crystallography and NMR* (New Haven, Yale University Press; 1992)]. The model for these calculations was a subset of the crystallographic dimer structure of the homologous protease from the VZV alpha-herpes virus (see Example 6 below). Each monomer in the search model was derived from a total of 177 amino acid residues from the VZV protease structure with the sidechains truncated to alanine. The residues of the core secondary structure of the VZV protease were included in the model: residues 11-22, 46-91, 95-124, 137-183 and 189-230 of SEQ ID NO: 5. The rotation function calculation was carried out with data between 15 and 4.0 Å resolution with a maximum search vector length of 38 Å. The top peak in the rotation function from this dimer model was 4.9 s. The translation function was calculated with data between 8 and 4 Å resolution. The top solution was at 9.3 s. Rigid body refinement of the two monomers reduced the R factor to 0.48 for all data to 3.5 Å.

#### E. Phase Combination

Using difference Fourier methods, phases derived from the molecular replacement solution were consistent with those generated from MIR. The combination of these two sets of phases using the program SIGMAA [Read, R. J. Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors *Acta Cryst.* A42, 140-149 (1986)] resulted in an overall figure of merit of 0.67. This was followed by one round of non-crystallographic symmetry averaging using the density modification program *dm* [Cowtan, K. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31, 34-38 (1994)] resulting in an improved overall figure of merit of 0.83. Non-crystallographic symmetry is symmetry that exists locally within the asymmetric unit of the crystal. This information can be used to produce averaged electron density maps in which noise will cancel out and therefore can be used as a phase restriction to improve phasing. The calculated electron density map following this procedure showed side chain density, derived solely from the MIR phases, that was very well-defined and easily interpretable.

#### F. Model Building and Refinement

30 The electron density allowed placement of almost all of the side chains in the original model using the program Xfit (McRee). Remaining effort was focused on building the missing 27% of the structure that was not part of the molecular replacement model. Two more rounds of density modification with the combined MIR phases allowed placement of an additional alpha helix, several loops and the DIP ligand in the active site. The electron density map resolution was extended to 2.5 Å resolution using *dm*, more residues were added and the model refined using X-PLOR.

When building the model, each of the amino acid residues was manually positioned in its electron density, allowing for a unique position for each atom in the DIP-liganded HSV-2 protease in which each position is defined by a unique set of atomic coordinates (X,Y,Z) as shown in Figure 2A-2F. Starting with these atomic coordinates, a diffraction pattern was calculated and compared to the experimental data. The difference between the calculated and experimentally determined diffraction patterns was monitored by the value of R-factors ( $R\text{-factor} = \frac{\sum |F_o| - |F_c|}{\sum F_o}$ ). The refinement (using XPLOR) of the structural model necessitates adjustments of atomic positions to minimize the R-factor, where a value of about 20% is typical for a good quality protein structure.

Cycles of model building with XTALVIEW and refinement with the computer program XPLOR produce a final model including 217 amino acids with 43 solvent molecules. Three segments of residues are found disordered in the crystal: 104-110, 134-140, and the first 16 residues of the N-terminus [SEQ ID NO: 4]. A total of 14605 reflections were included in the final refinement (10.0 - 2.5 Å), giving an R-factor ( $\frac{\sum |F_o| - |F_c|}{\sum F_o}$ ) of 20.5%. The rms bond length is 0.016 Å and rms bond angle is 1.9°. The program PROCHECK [R. A. Laskowski *et al.*, *J. Appl. Crystallogr.* 26: 283-291 (1993)] was used to check the stereochemical and geometrical outliers in the final structure, and the result is very satisfactory.

50 The statistics of structure determination are reported in Fig. 42, where  $R_{\text{sym}} = \frac{\sum |I - \langle I \rangle|}{\sum I}$ ,  $I$  is the observed intensity and  $\langle I \rangle$  is the average intensity of multiple observations;  $R_{\text{iso}} = \frac{\sum |F_{\text{PH}} - F_P|}{\sum F_P}$ ; Phasing power = rms isomorphous difference/ rms residual lack of closure;  $R_{\text{cullis}} = \frac{\sum |F_H - F_c|}{\sum |F_H|}$ ,  $F_H$  and  $F_c$  are the observed and calculated heavy atom structure factor amplitudes for centric reflections;  $R\text{-factor} = \frac{\sum |F_o| - |F_c|}{\sum F_o}$ . XPLOR refinement was performed according to A. F. Brünger *et al.*, *Science*, 235: 458-460 (1987), from 10-2.5 Å. The number of reflections used ( $F > 2\sigma$ ) = 14605, the R-factor was 20.5%, the number of protein atoms (non-H) was 3364; the number of solvent atoms was 43; the RMS bond length was 0.016 Å; and the RMS bond angles = 1.919°. Mean coordinate error (0.3 Å) was performed according to P. V. Luzatti, *Acta Cryst.* 5: 802-810 (1952). MIR overall mean figure of merit (15-3.0 Å) = 0.62 Å; overall figure of merit after phase combination = 0.67 Å; mean figure of merit following density modification 100-3.0 Å = 0.838 Å.

Using the final atomic coordinates (Figs. 2A-F) one can calculate distances between a pair of atoms, angles

between any three atoms, and dihedral angles between any four atoms, such as listed in Figs. 8A-X, 11A-LLL and 14.

The unliganded HSV-2 protease structure was solved using difference Fourier methods using the refined CIP-liganded HSV-2 protease structure. Since the cell dimensions and space group of the unliganded and liganded (DIP) structure were the same, the DIP-liganded HSV-2 protease structure could be used directly to determine the phases of the unliganded structure, without using heavy atom derivatives or molecular replacement. The unliganded HSV-2 protease model coordinates could then be determined and refined as described for the DIP-liganded HSV-2 protease structure. Three segments of residues were disordered in the unliganded HSV-2 protease structure: 1-16, 104-112, and 134-140 [SEQ ID NO: 4]. A total of 10127 reflections were included in the final refinement (7.0-2.8 Å) giving an R-factor ( $\sum |F_o| - |F_c| / \sum F_o$ ) of 22.4%. The rms bond length is 0.017 Å and rms bond angle is 2.1°. The program PROCHECK [R. A. Laskowski *et al.*, *J. Appl. Crystallogr.*, 26: 283-291 (1993)] as used to check the stereochemical and geometrical outliers in the final structure, and the result is very satisfactory.

#### Example 2: Analysis of the Structure of the HSV-1 Protease

The HSV-1 protease (see Fig. 1, SEQ ID NO: 3) was cloned, expressed and purified as follows:

##### A. Expression, Purification and Crystallization

HSV-1 protease was expressed and purified as described above in Example 1 for HSV-2. HSV-1 was crystallized in 45 mM Tris buffer pH 8.5, 88 mM MgCl<sub>2</sub> and 8.8% PEG 8000 at 4°C.

##### B. X-ray Diffraction Characterization

A HSV-1 protease crystal was subjected to X-ray diffraction using the techniques described in Example 1 above. The crystal diffracted to 3.5 Å resolution. By registering the position and intensity of many tens of thousands diffraction spots using the computer program XDS [Kabsch, W. *J. Appl. Cryst.* 21, 916-924 (1988)], the crystal has been determined to be the orthorhombic space group P1, with  $a = 79.62$  Å  $b = 81.18$  Å and  $c = 93.36$  Å,  $\alpha = 115.49^\circ$   $\beta = 98.36^\circ$   $\gamma = 109.18^\circ$ . The native data is 78.4% complete to 3.5 Å with an  $R_{\text{sym}}$  ( $\sum |I - \langle I \rangle| / \sum \langle I \rangle$ ) of 0.059. By established methods, an asymmetric unit was calculated to have either six or eight protein molecules (three or four dimers).

##### C. Molecular Replacement

The HSV-1 protease structure was solved by the method of molecular replacement using the program AMoRe in the CCP4 Suite [Collaborative Computational Project, Number 4 *Acta Crystallogr.* D50, 760-763 (1994)]. The model for these calculations was the entire known crystallographic dimer structure of the highly homologous protease from the HSV-2 alpha-herpes virus complexed with the covalently bound inhibitor DFP. The model was defined as residues 17-103, 111-133, 141-247 [SEQ ID NO: 3] for each monomer with no inhibitor atoms included. The rotation function calculation was carried out with data between 8 and 4.0 Å resolution with a maximum search vector length of 31.3 Å. Only three pairs of peaks were found with peak height greater than 0.5 (maximum peak height). The pairs of peaks reflected the non-crystallographic symmetry of the dimer. Using data between 8 and 4 Å resolution, the translation function was calculated by fixing the top solution in the P1 cell and searching for a second molecule. This yielded a peak with a correlation coefficient of 41.0% and an R-factor of 38.5%. The top two solutions were then fixed to search for a third, yielding a peak of comparable height to the second solution. A search for a fourth solution showed smaller peaks of all about the same height. To see if there were three or four dimers in the cell a rigid body fit was run to search for four dimers using the top three solutions with several of the similar peaks generated in the last translation function output. The fitting function yielded four peaks with a correlation coefficient of 45.6% and R-factor of 37.1%. Alternatively, when the top three dimer solutions were fit, the correlation coefficient rose to 55.9% with an R-factor of 33.1%. Several other combinations of peaks were tried as controls and none yielded satisfactory results as compared to the top three peaks. A packing diagram was determined and visually inspected using the program Xfit, [McRee, D.E. *Practical Protein Crystallography*, (San Diego, Academic Press, 1993)] showing no overlaps between symmetry related molecules.

##### D. Model Placement and Refinement

When the molecular replacement solution was placed into the P1 cell, the model was changed slightly in which residues that were different between the two sequences were truncated to alanine to reduce the bias of the HSV-2 protease phases in the calculation of the electron density map. Unfortunately, most of the significant differences between HSV-1 protease and HSV-2 protease in the sequence are present in regions missing in the HSV-2 protease structure [N-terminus and 134-140 of SEQ ID NO: 4], so that many of the changes in the new model, limited by the 3.5 Å resolution, would not be reflected in the electron density map. As a control, to ensure that the map revealed the contributions

of the HSV-1 data, six phenylalanine or tyrosine residues were truncated to alanine on each monomer. Fourier coefficients were calculated using the program SIGMAA [Read, cited above (1986)]. This was followed by phase improvement by non-crystallographic symmetry averaging using the program dm [Cowtan, cited above (1994)]. Non-crystallographic symmetry is symmetry that exists locally within the asymmetric unit of the crystal. This information can be used to produce averaged electron density maps in which noise will cancel out and therefore can be used as a phase restriction to improve phasing.

The calculated electron density map following this procedure showed side chain density that reflected the HSV-1 sequence, within the limits of the resolution, and clearly showed density for the phenylalanine and tyrosine side chains that were omitted from the model, indicating the electron density did reflect the contributions from the HSV-1 protease data.

The residues unique to the HSV-1 protease sequence were built into the model using Xfit (McRee). When building and changing the model, each of the amino acid residues was manually positioned in its electron density, allowing for a unique position for each atom in the HSV-1 protease in which each position is defined by a unique set of atomic coordinates (X, Y, Z) as shown in Figs. 6A-B. Starting with these atomic coordinates, a diffraction pattern was calculated and compared to the experimental data. The difference between the calculated and experimentally determined diffraction patterns was monitored by the value of the R-factor ( $R\text{-factor} = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}$ ). The refinement of the structure was done by rigid body where the fit of the model could be refined by rotation and translation of the entire model. Further positional refinement was not possible because of lack of experimental data as compared to refinement parameters.

The final model has 214 amino acids. Three segments of residues are found disordered in the crystal: 102-110, 134-143, and the first 14 residues of the N-terminus of SEQ ID NO: 3. A total of 12346 reflections were included in the final refinement (10-3.5 Å), giving an R-factor ( $\frac{\sum |F_o| - |F_c|}{\sum |F_o|}$ ) of 36.9%.

Using the final atomic coordinates (Fig. 6) one can calculate distances between a pair of atoms, angles between any three atoms, and dihedral angles between any four atoms, such as listed in Figs. 10A-B, 13A-D and 16.

### Example 3: Analysis of the Structure of the CMV Protease

The CMV protease (see Fig. 1, SEQ ID NO: 1) was cloned, expressed and purified as follows:

#### A. Expression, Purification and Crystallization

CMV A143V protease was expressed and purified as described for HSV-2 and HSV-1. After screening against about a thousand different conditions, the protein was finally crystallized in 30% PEG400 at pH4. Large crystals are approximately 0.4mm x 0.3mm x 0.3mm in size.

#### B. X-ray Diffraction Characterization

The CMV protease crystal was subjected to x-ray diffraction using the techniques described above for HSV-2 and HSV-1 protease crystals, with the exception that the anode machine was operated at 50 KV x 100 mA electrode power. The crystal diffracted to 3.0 Å resolution. By registering the position and intensity of many tens of thousands diffraction spots using the computer program XENGEN, the crystal has been determined to be tetragonal crystal system and  $P4_322$  space group. The unit cell dimensions are  $a=b=58.7$  Å and  $c=131.0$  Å. By established methods, an asymmetric unit was calculated to have one protein molecule. The crystal contains an estimated 40% solvent.

A higher resolution diffraction data set (2.5 Å) was collected at the Cornell Synchrotron Laboratory (CHESS) A-1 beamline using a CCD detector. The data was processed with the programs DENZO/SCALEPACK [Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. Bailey, S.) 56-62, Daresbury Laboratory, Warrington (1993)]. Others were collected with a Siemens multiwire detector on a Siemens  $\text{CuK}_\alpha$  source and processed with XENGEN [A. J. Howard *et al.*, *J. Appl. Crystollogr.* A47: 110-119 (1994)].

#### C. Heavy Atom Derivatives:

Using the MIR methods described in Example 1, by testing various heavy metal compounds, the useful derivatives were prepared by soaking the native crystals with saturated  $\text{MeHgCl}$  (at pH4 or 5), saturated Baker's Dimercury, 1mM  $\text{UO}_2\text{Ac}_2$ , 1mM  $\text{K}_2\text{PtCl}_4$ , 0.5mM  $\text{LuCl}_3$  or  $\text{SmCl}_3$  for one to four days. The X-ray diffraction data of each of the derivatives were then collected by the same methods described above. Data collection statistics for native and heavy atom derivatives are shown in Fig. 43. Heavy atom positions were identified by difference Patterson and difference Fourier methods using the programs in the CCP4 suite [Collaborative Computational Project, Number 4 *Acta Crystallogr.* D50, 760-763 (1994)]. Anomalous signals from three of the derivatives allowed the determination of the chirality of space group and heavy atom coordinates. Heavy atom refinement and phasing were carried out using the program MLPHARE [Z. Otwinowski, cited above (1991)]. Using the initial phases obtained through the MIR methods, a map of electron density

within the crystal unit cell was calculated. Because electrons are heavily distributed in the immediate vicinity of the centers of atoms, the positions of protein atoms are registered according to the electron density map. The clarity of the electron density map was improved with the methods of solvent flattening, histogram matching and skeletonization.

#### 5 D. Model Building and Refinement

Using the three-dimensional electron density map obtained from above experiments, the polypeptide chain of the CMV protease can be traced without ambiguity. 193 residues (most with side chains) were built using the 3-D computer graphics program XTALVIEW [McRee, D.E., cited above (1993)]. XTALVIEW was used in building models of the CMV protease structure. Each of these 193 amino acids residues was manually positioned in its electron density, allowing for a unique position for each atom in the CMV protease in which each position is defined by a unique set of atomic coordinates (X,Y,Z) as shown in Figs. 17A-E. Starting with these atomic coordinates, a diffraction pattern was calculated and compared to the experimental data. The difference between the calculated and experimentally determined diffraction patterns was monitored by the value of R-factors ( $R\text{-factor} = \frac{\sum |F_o| - |F_c|}{\sum F_o}$ ). The refinement (using XPLOR) of the structural model necessitates adjustments of atomic positions to minimize the R-factor, where a value of below 20% is typical for a good quality protein structure and a value of higher than 20% usually indicates the need of further refinement.

The initial model of CMV protease contains about 70% of the amino acids, having a starting R-factor of 43.8% using the diffraction data from 10 to 3.0 Å. The computer program XPLOR was used to carry out the refinements, and the models were improved gradually after many iteration cycles. The R-factor was decreased to 28.3% after 200 cycles of positional refinement with XPLOR. The final R-factor is 18.7% the CMV protease structure. The program PROCHECK [R. A. Laskowski *et al.*, cited above (1993)] was used to check the stereochemical and geometrical outliers in the final structure, and the result is very satisfactory.

The statistics of structure determination data is reported in Fig. 43, where  $R_m$ ,  $R_{iso}$ , and  $R_{Cullis}$ . R-factor are as defined in Example 1 above.  $R_c(\text{ano})$  is defined for anomalous amplitudes of non-centric reflections similar to the original  $R_{Cullis}$  formula. As described in Example 1, XPLOR refinement was performed according to A. T. Brunger *et al.*, cited above. More particularly, resolution included: 7.0-2.5Å. No. reflections used (>1σ):7193; R-factor: 0.185, No. protein atoms (non-H):1604 (202 aa); No. of solvent atoms (non-H):73; MIR as figure of merit (30-3.2Å):0.70. Mean coordinates error: 0.4Å; RMS bond length: 0.017Å; RMS bond angle (2.2 degrees). Mean coordinates error was performed according to the SIGMAA program [R. Read, *J. Appl. Crystallogr.*, A42: 140-149 (1986)].

Using the final atomic coordinates (Figs. 17A-E) one can calculate distances between a pair of atoms, angles between any three atoms, and dihedral angles between any four atoms, such as listed in Figs. 18A-C, 19A-D and 20.

#### 35 Example 4: Cloning and Expression of the VZV Protease

The VZV protease gene was located in the complete VZV genome by homology to the protease genes from the HSV-1 and CMV herpes viruses. The VZV genomic sequence used for this analysis was as published by A. Davison & J. Scott, *J. Gen. Virol.*, 67:1759-1816 (1986). The open reading frame for the protease/capsid-encoding gene (equivalent to the HSV-1 UL26 gene) was found to start at base 62,138 and stop at base 60,324 encoding 605 amino acid residues. This open reading frame had been referred to as gene 33 in the above mentioned publication.

The 236 amino acid long protease catalytic domain [SEQ ID NO: 5] was located by identifying the R site that defines the carboxyl-terminal end of all known herpes virus proteases. An alignment of such known R sites is shown in Table I. These cleavage sites are highly conserved (as shown by underlined residues) with cleavage occurring between alanine and serine residues, as indicated by \*\*\*.

TABLE I

Protease	R Site Sequence	Sequence ID No:
HSV-1	Tyr-Leu-Gln-Ala* <u>Ser</u>	3
HSV-2	Tyr-Leu-Gln-Ala* <u>Ser</u>	4
CMV	Tyr-Val-Lys-Ala* <u>Ser</u>	1
EBV	Tyr-Leu-Lys-Ala* <u>Ser</u>	6
VZV	Tyr-Leu-Gln-Ala* <u>Ser</u>	5

### A. Design of VZV Protease synthetic gene

In an effort to optimize bacterial expression of the VZV protease catalytic domain, a synthetic gene was constructed using codons that are found in proteins highly expressed in *E. coli*. In addition, a number of constructs were made with the goal of facilitating purification of the protein as an active enzyme. Most of the constructs were aimed at producing the authentic species believed to be made during viral infection.

The synthetic gene was designed as follows: A 788 bp VZV protease gene fragment was designed with an NcoI restriction site at the 5' end and XbaI site at the 3' end. These restriction sites are useful for subsequent cloning of the gene fragment in a suitable expression vector. A unique BstE2 restriction site was introduced in the middle of the gene fragment without altering the amino acid sequence. This restriction site was later used to ligate the two synthetic fragments together. It was decided to construct this gene in two portions for the ease of gene synthesis.

The 5' portion of the gene was about 370 bp long and the 3' portion was about 418 bp.

### B. Design and synthesis of oligonucleotides

Four megaprimers (primers which are more than 100 bases long) with about 25 bp overlapping ends were designed using the Oligo 4.0 software from National Biosciences Inc. Care was taken to avoid mismatching of overlapping ends. All primers were synthesized on an Applied Biosystem DNA Synthesizer (Model 394) using 40 nM polystyrene columns. Crude oligonucleotide primers were used to assemble the gene fragments. For each portion of the gene, two PCR primers containing unique restriction sites were made using the same DNA synthesizer. These oligonucleotides were referred to as 'nested primers'.

### C. Gene Synthesis

The gene synthesis was carried out using the procedure described by Rosen et al, *BioTechniques*, 9(3) (1990) with some modifications. For each portion of the gene, two megaprimers oligonucleotides were phosphorylated using the standard kinase procedure [Sambrook et al, *Molecular Cloning. A Laboratory Manual*, 2nd edit., Cold Spring Laboratory, New York (1989)]. In the first polymerase chain reaction (PCR), 0.5 to 1 µg of each of the four megaprimers were mixed together and the PCR was carried out using dNTPs and a mixture of Taq and Vent DNA polymerases (6:1 v/v). Only about 15 cycles of PCR were carried out using the Perkin Elmer PCR 9600 thermocycler (94°C for 30 seconds, 52°C for 30 seconds, 72°C for 45 seconds).

The product of this PCR reaction was used as a template in the second PCR reaction along with 5' and 3' gene specific primers containing unique restriction sites ('nested primers'). PCR reactions was carried out for 25-30 cycles using similar cycle times as above.

About 10 µl of reaction product was analyzed on a 1% agarose gel. PCR products showing a correct size band were then subcloned in the PCR II vector [Invitrogen, San Diego, CA]. The DNA sequence of the synthetic fragments was confirmed by automated DNA sequencing.

### D. VZV protease constructs

Six constructs were prepared for the expression of the VZV protease catalytic domain and are illustrated as Figs 45A to 45E [SEQ ID NOS: 7 through 11, respectively]. Authentic VZV protease [SEQ ID NO: 5] contained a protease domain authentic at both amino and carboxyl termini.

H6(N)VZV protease [SEQ ID NO: 7] contained an authentic protease domain preceded at the amino-terminus by six histidine residues (underlined) followed by an enterokinase cleavage site (bold, underlined). The amino-terminal sequence of this construct is: MGHHHHHHSSGHI**DDDDK**-MAAE...

LQA-H6(C) VZV protease [SEQ ID NO: 8] contained an authentic protease domain followed by six histidine residues.

LQAS-H6(C) VZV protease [SEQ ID NO: 9] contained an authentic protease domain followed by a serine residue and six histidine residues (underlined).

LQAS-12aa ext H6(C) VZV protease [SEQ ID NO: 10] contains an authentic protease domain followed by a serine residue, 12 residues normally found after the LQAS R-site (bold underlined) and six histidine residues (underlined). The carboxyl-terminal sequence of this construct is: ...LQAS-**TGYGLARITNVN**-HHHHHH.

Delta9 LQAS- 12 aa ext H6(C) VZV protease [SEQ ID NO: 11] contained an authentic protease domain deleted at the amino-terminus (first nine natural residues removed and Cys 10 replaced by Met) and followed at the carboxyl-terminus by a serine residue, 12 residues normally found after the LQAS R-site (bold underlined) and six histidine residues (underlined). The amino-terminus of this construct is MEALYV ...; and the carboxyl-terminal sequence of this construct is: ...LQAS-**TGYGLARITNVN**-HHHHHH.

**E. Expression of VZV protease constructs**

All constructs were inserted in the *E. coli* expression vector pET16b (Novagen, Madison, WI) in which the inserted gene is under the control of the inducible T7 promoter. The vectors were introduced in the BL21(DE3) *E. coli* strain (Novagen) by standard transformation techniques. The transformed cells were grown to OD<sub>650</sub>=0.5 and then treated with IPTG at 10 mM to induce expression from the T7 promoter. The cells were then aerated for an additional 2 hours and collected by centrifugation. Cell extracts were analyzed for expression by SDS-PAGE followed by Coomassie staining or western blot analysis using a polyclonal antibody against HSV-1 protease, called Anti-95370.

Anti-95370 is a rabbit anti-HSV-1 protease polyclonal antiserum prepared by fusing the complete HSV1 UL26 gene to the C-terminus of a truncated GalK gene in the pOTSKF33 vector described in C. S. Chiang et al. *Clin. Chem.* 35(6):946-952 (1989). The fusion protein was expressed by conventional protocols and after cell lysis, the insoluble fraction was gel purified using preparative SDS-PAGE. The fusion protein was electroeluted and used to immunize rabbits by standard protocols. The resulting Anti-95370 antisera was shown to cross-react with the VZV protease.

**Example 5: Purification of VZV protease constructs**

No purification work was done on the authentic VZV protease construct. The other constructs were purified as follows:

**A. Purification of H6(N)NZV Protease [SEQ ID NO: 7]**

Expression of this protease construct was examined by comparing two hour inductions at 25°C and 37°C. Cells (2-3g) were resuspended in 50 mM Tris pH 8.0, 300 mM NaCl at a ratio of 10 ml/g cells and lysed by sonication on ice. Subsequent purification procedures were performed at 4°C. After centrifugation at 30,000xg, the soluble fraction was further purified by a one hour batch incubation with NiNTA agarose (Qiagen) followed by column chromatography with imidazole washes and elution. Samples were analyzed by Coomassie stained SDS-PAGE and Western blot using Anti-95370 polyclonal antibody described above.

More protease was expressed at 37°C but the majority of the protease was insoluble under both conditions. The soluble protease appeared divided between full length and truncated (identified as C-terminal des30) forms. The majority of the product eluted at 50 mM imidazole rather than the expected 250 mM. The 50 mM eluate was 90% pure, ~90% truncated, active against the JM82 peptide substrate (Ac-HTYLQA\*SEKFKMWG; \* represents the cleavage site) [SEQ ID NO: 16] and had the correct N-terminal sequence. The activity was attributed to full length product.

**B. Purification of LQA-H6(C) VZV Protease [SEQ ID NO: 8] and LQAS-H6(C) VZV Protease [SEQ ID NO: 9]**

Cells (5-10g) induced in shake flasks at 25°C and 37°C expressing these constructs were resuspended in 50 mM Tris pH 8.0, 300 mM NaCl at a ratio of 10 ml/g cells and lysed with the Avestin homogenizer (~12,000 psi). The lysate was centrifuged and the soluble fraction was chromatographed on NiNTA agarose. More protease eluted with 50 mM imidazole than with 250 mM imidazole for both constructs induced at 37°C. The majority of product eluted at 250 mM imidazole for LQA-H6(C) VZV protease induced at 25°C. The relative elutions were consistent in RP-HPLC, Coomassie and Western analyses. All products appeared to be of equivalent size.

The products from both constructs had the predicted N-terminal (desMet). The 50 mM eluate from LQAS-H6(C) VZV protease [SEQ ID NO: 9] was concentrated to ~0.6 mg/ml, made 10 mM DTT, 1 mM EDTA and 10% glycerol and incubated at 4°C. Slight activity (<10% specific activity of fully processed protease) against the JM82 peptide substrate was detected 10 days later and confirmed after another 10 days. When the shake flask preparation was repeated and an additional Superdex 75 chromatography step was added, the final product still had only about 10% of the potential activity. Scale-up of LQAS-H6(C) VZV protease [SEQ ID NO: 9] using cells grown in a 10 liter fermentor and induced at OD 0.7 or OD 5.0 at 37°C for 1 hour was unsuccessful.

Coomassie-stained SDS-PAGE gels indicated that lysis was successful but no protease could be detected in NiNTA eluates using RP-HPLC. Western blot detected the best expression with cells induced at OD 0.7 but the majority of the product was insoluble. The same levels of product were detected in the NiNTA agarose load and unbound for the 10L fermented cells suggesting that product did not bind. Product was detected in the load from cells grown in shake flask but not in the unbound fraction suggesting complete capture by NiNTA agarose. The product from all samples had the same apparent molecular weight. The reason for the failure to bind to NiNTA agarose is unknown.

**D. Purification of LQAS-12aa ext H6(C) VZV Protease [SEQ ID NO: 10] and delta9 LQAS- 12 aa ext H6(C) VZV Protease [SEQ ID NO: 11]**

300g of cells from LQAS-12aa ext H6(C) VZV protease [SEQ ID NO: 10] were resuspended in buffer A (50 mM Tris,

pH8.0, 300 mM NaCl) to a final volume of 3L and the cells were lysed with an Avestin homogenizer at ~12,000 psi. The homogenate was centrifuged for 1 hour, 4°C at 14,000 rpm (30,000 g). The supernatant was collected and added to NiNTA agarose (Qiagen) at a ratio of 1 ml/10g cells. After incubation with rotation at 4°C for 1 hour, the resin was collected by centrifugation for 10 minutes at 3000 rpm. The supernatant was removed with a peristaltic pump and the resin was packed into a Pharmacia 2.6 cm XK column with a 25 ml pipette. The column was washed to baseline absorbance (0.5 mV) with buffer A at 2.5 ml/minute. After washing with 20% B (50 mM imidazole; buffer B is 50 mM Tris, pH8.0, 300 mM NaCl, 250 mM imidazole), the protease was eluted with 100% B (250 mM imidazole). Glycerol was added to the 250 mM eluate to 10%, DTT to 10 mM, and EDTA to 1 mM final concentrations. The sample was filtered with a STE-RIVEX 0.45mm (Millipore) filter and transferred to an Amicon 50 ml stirred cell. After concentration to 5 ml, the sample was diluted to 50 ml with SEC buffer (25 mM HEPES, pH8.0, 50 mM NaCl, 1 mM EDTA, 5 mM DTT) and concentrated to <5 ml. After filtration the sample was stored overnight at 4°C.

The sample was chromatographed on a 2.6 x 60 cm Superdex column equilibrated in SEC buffer. Protease fractions determined by absorbance at 280 nm were pooled and concentrated in a 10 ml stirred cell. Precipitate was removed by centrifugation and the sample was filtered (Millipore ULTRAFREE MC 0.22µm). The VZV protease product was concentrated to 2 mg/ml and diluted with an equal volume of glycerol before storage at -20°C for protease assays. Alternately, protease was concentrated to ~10 mg/ml for crystallography.

The protease product from LQAS-12aa ext H6(C) VZV [SEQ ID NO: 10] was determined to be of the predicted mass by MALDI-MS without the N-terminal Met and with a carboxyl terminus corresponding to authentic protease (i.e., ending with LQA as a result of auto-processing of the 12 amino acid extension at the R site). In fact, it was determined that this construct, which contained a carboxy tail with a hexahistidine tail for binding to NiNTA, following 13 additional amino acids after the LQA mature protein carboxy terminal amino acids, permitted the production of a properly cleaved carboxy terminus for the protein. The protease from delta9 LQAS- 12 aa ext H6(C) VZV [SEQ ID NO: 11] had the predicted mass while retaining the N-terminal Met and also ending with LQA. Both proteases were active against the JM82 substrate. Purified protease was a single peak on RP-HPLC and SEC and 95% pure on RP-HPLC.

Protease from delta9 LQAS- 12 aa ext H6(C) VZV [SEQ ID NO: 11] was primarily prepared for crystallography (to eliminate structural disorder at the amino terminus) with a typical yield of 4 mg/300 g *E. coli* cells. Protease from LQAS-12aa ext H6(C) VZV [SEQ ID NO: 10] yielded as much as 17 mg/300 g *E. coli* cells.

These modified VZV protein constructs are useful in the crystallization of VZV protease as described below in Example 6, as well as for other biophysical structural studies of VZV protease. The constructs are also useful in biochemical assays to identify compounds which inhibit and/or interact with VZV protease (see Examples 7 and 8 below).

#### Example 6: Crystallization of the VZV Protease

Protease from delta9 LQAS- 12 aa ext H6(C) VZV [SEQ ID NO: 11] was crystallized in 0.1 M phosphate buffer pH6.2 and 2.5 M NaCl. Large crystals are approximately 0.5mm x 0.2mm x 0.2mm in size.

#### **A. X-ray Diffraction Characterization**

A VZV protease crystal was subjected to x-ray diffraction characterization, using essentially the same methods as described for CMV. The crystal diffracted to 3.0 Å resolution. By registering the position and intensity of many tens of thousands diffraction spots using the computer program XENGEN, the crystal has been determined to be the hexagonal space group P6<sub>4</sub>22, with a=b=90.0 Å and c=117.4 Å. By established methods, an asymmetric unit was calculated to have one protein molecule. The crystal contains an estimated 60% solvent. The native data is 90% complete to 3.0 Å with an  $R_{\text{merge}}$  ( $\sum |I - \langle I \rangle| / \sum I$ ) of 0.07.

#### **B. Heavy Atom Derivative**

Single isomorphous replacement (SIR) methods were used. The useful derivatives were prepared by soaking the native crystals with 1mM KPt(CN)<sub>2</sub> for one day. Using the initial phases obtained through the SIR methods, a map of electron density within the crystal unit cell could be calculated. Because electrons are heavily distributed in the immediate vicinity of the centers of atoms, the positions of protein atoms are registered according to the electron density map. The clarity of the electron density map could be improved with the methods of solvent flattening, histogram matching and skeletonization. The derivative data is 81% complete to 4.5 Å with an  $R_{\text{merge}}$  of 0.14 and  $R_{\text{iso}}$  ( $\sum |F_{\text{PH}} - F_{\text{P}}| / \sum F_{\text{P}}$ ) of 0.19. This derivative gave a phasing power of 1.2 and  $R_{\text{Cullis}}$  ( $\sum |F_{\text{H}_0} - F_{\text{H}_c}| / \sum |F_{\text{H}_0}|$ ) of 0.79. In this case, the phase information from SIR is not sufficient for structure solution.

#### **C. Molecular Replacement**

A molecular replacement (MR) solution using XPLOR [A. Brunger et al. Science 235, 458-460 (1987)] was suc-

cessfully identified only after all the less conserved regions were excluded in the search model which originated from the CMV protease structure. The rotation solution (8.0-4.0 Å) is the highest peak and is 25s above the mean and 1s higher than the second highest peak. Translation searches were carried out in two possible spacegroups P6<sub>2</sub>22 and P6<sub>4</sub>22, and the latter gave a better solution of 5s in peak height and 52.6% in R-factor (8.0 - 3.0 Å). After rigid body refinement using 8.0-3.0 Å data, the R-factor dropped to 50.6%. When examining the crystal packing, a tight dimer interface was found that corresponds to the same interface in CMV protease. Using the calculated phases from the molecular replacement solution, a heavy atom position was identified using difference Fourier methods that is identical to the one found using difference Patterson methods. This position is in the vicinity of Cys157, which further confirmed the correctness of these results.

#### D. Phase Combination

The crystal structure of VZV protease was determined using the combination of single isomorphous replacement (SIR) and molecular replacement methods. Neither the SIR map nor the MR map seemed to be interpretable. Combining the phases from both sources, the overall figure of merit was only 0.39 and the map is still quite noisy. Fortunately, there is 60% of solvent in the crystal. After solvent flattening and histogram matching, the electron density map became very clear.

#### E. Model Building and Refinement

The polypeptide chain of the VZV protease [SEQ ID NO: 5] can be traced without ambiguity using the three-dimensional electron density map obtained from the above-experiments and the methods described in Example 3 for CMV above, with the exception that 211 residues (most with side chains) were built.

Cycles of model building with XTALVIEW program and refinement with the XPLOR computer program produce a final model including 211 amino acids without any solvent molecules. Fifteen residues are found disordered in the crystal: 127-136 and 232-236 of SEQ ID NO: 5. A total of 4903 reflections were included in the final refinement (7.0 - 3.0 Å), giving an R-factor ( $\sum |F_o| - |F_c| / \sum F_o$ ) of 22.3% without refining temperature factors. The rms bond length is 0.014 Å and rms bond angle is 2.1°. His 52 and His 139 were refined as carrying a single proton at the ND1 atom. The program PROCHECK [R. A. Laskowski et al., *J. Appl. Crystallogr.*, 26: 283-291 (1993)] was used to check the stereochemical and geometrical outliers in the final structure, and the result is very satisfactory.

Using the final atomic coordinates (Figs. 22A-C) one can calculate distances between a pair of atoms, angles between any three atoms, and dihedral angles between any four atoms, such as listed in Figs. 24-26.

#### EXAMPLE 7: Protease Activity Assays

The biological function of the HSV-2, HSV-1, CMV, and VZV proteases *in vivo* is to specifically cleave the Ala-Ser peptide bonds within a large protein substrate molecule. For routine *in vitro* assay of the protease activity, use of a large protein substrate is inconvenient and very expensive.

##### A. HSV-2 Protease

For HSV-2 protease, a small peptide substrate having the sequence dabsyl-DNAVEA\*SSKAPLK-(dansyl-II)-OH entitled FQ7 (based on VZV m site) [SEQ ID NO: 17] has been synthesized in place of large protein substrate. In the presence of the HSV-2 protease, the FQ7 peptide will be cleaved at the A\*S peptide bond, and the product will be the two halves of the substrate. The peptide has been designed so that the cleaved molecules will give rise to strong fluorescence signals.

Therefore, the enzymatic activity of HSV-2 protease can be measured quantitatively by the intensity of the fluorescence signal. This is a very sensitive assay method called fluorescence quenching (FQ). For instance, see, "Principles of Fluorescence Spectroscopy", Lakowicz, J. R., Plenum Press, N.Y. 1983.

In experiments conducted on the HSV-2 protease, the optimized assay conditions call for the use of 520 nM of the HSV-2 protease, 30% of sucrose and 0.8 M citrate. In the presence of added inhibitors, the decreased amount of activity also quantifies the potency of the inhibitors.

##### B. HSV-1 Protease

For HSV-1 protease, a small peptide substrate having the sequence Ac-HTYLQA\*SEKFKMWG entitled JM82 [SEQ ID NO: 16] has been synthesized in place of large protein substrate. In the presence of the HSV-1 protease, the JM82 peptide will be cleaved at the A\*S peptide bond, and the product will be the two halves of the substrate. Activity was measured by quantification of the two halves of the substrate using HPLC.

In experiments conducted on the HSV-1 protease, the optimized assay conditions call for the use of 0.3 mg/ml of the HSV-1 protease, 30% of sucrose and 0.8 M citrate in a buffer of 25 mM Hepes (pH 8.0), 50 mM NaCl, 10mM DTT, 1 mM EDTA and 10% Glycerol. In the presence of added inhibitors, the decreased amount of activity also quantifies the potency of the inhibitors.

### C. CMV Protease

For CMV protease, a small peptide substrate having the sequence Dbs-RGVVNASSRLAKK-DNS(II) entitled FQ8 [SEQ ID NO: 18] has been synthesized in place of large protein substrate. In the presence of the CMV protease, the FQ8 peptide will be cleaved at the A\*S peptide bond, and the product will be the two halves of the substrate. As for HSV-1 and HSV-2 proteases, the peptide substrate has been designed so that the cleaved molecules will give rise to strong fluorescence signals.

In experiments conducted on the CMV protease, the optimised assay conditions call for the use of 20 mM of the CMV protease and 30% of sucrose. In the presence of added inhibitors, the decreased amount of activity also quantifies the potency of the inhibitors.

### D. VZV Protease

This assay was performed as described for the proteases above, making use of the FQ7 small peptide substrate. In experiments conducted on the VZV protease, the optimized assay conditions call for the use of 20 nM of the VZV protease, with buffer of 50 mM Hepes, pH8, 150 mM NaCl, 1 mM EDTA, 0.01% PEG with 0.8M citrate/30% sucrose. In the presence of added inhibitors, the decreased amount of activity also quantifies the potency of the inhibitors.

### Example 8: Method of Detecting Inhibitors

The three dimensional atomic structure can be readily used as a template for selecting potent inhibitors. Various computer programs and databases are available for the purpose. A good inhibitor should at least have excellent steric and electrostatic complementarity to the target, a fair amount of hydrophobic surface buried and sufficient conformational rigidity to minimize entropy loss upon binding.

There are generally several steps in employing the 3D structure as a template.

First, a target region is defined. In defining a region to target, one can choose the active site cavity of the herpes protease, or any place that is essential to the protease activity. As described above, for HSV-2, HSV-1, CMV and VZV proteases, the crystal structure is determined and therefore spatial and chemical properties of the target region are known.

Second, a small molecule is docked onto the target using one of a variety of methods. Computer databases of three-dimensional structures are available for screening millions of small molecular compounds. A negative image of these compounds is calculated and used to match the shape of the target cavity. The profiles of hydrogen bond donor-acceptor and lipophilic points of these compounds are also used to complement those of the target. One skilled in the art can readily identify many small molecules or fragments as hits.

Third, one may link and extend recognition fragments. Using the hits identified by above procedure, one can incorporate different functional groups or small molecules into a single, larger molecule. The resulting molecule is likely to be more potent and have higher specificity than a single hit. It is also possible to try to improve the "seed" inhibitor by adding more atoms or fragments that will interact with the target protein. The originally defined target region can be readily expanded to allow further necessary extension.

A limited number of promising compounds is selected via this process. The compounds are synthesized and assayed for their inhibitory properties. The success rate is sometimes as high as 20%, and it may still be higher with the rapid progresses in computing methods.

This invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

The disclosures of the patents, patent applications and publications cited herein are incorporated by reference in their entireties.

(2) INFORMATION FOR SEQ ID NO:17:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

15 Asp Asn Ala Val Glu Ala Ser Ser Lys Ala Pro Leu Lys  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:18:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site  
 (B) LOCATION: 1  
 30 (D) OTHER INFORMATION: /note= "Arg at amino acid position  
 1 is modified to contain a dabsyl group"

(ix) FEATURE:

(A) NAME/KEY: Modified-site  
 (B) LOCATION: 13  
 35 (D) OTHER INFORMATION: /note= "Lys at amino acid position  
 13 is modified to contain a dansyl-II g..."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

40 Arg Gly Val Val Asn Ala Ser Ser Arg Leu Ala Lys Lys  
 1 5 10

45

Claims

1. A composition comprising a herpes virus protease in crystalline form.
2. The composition according to claim 1 wherein said protease has an active site cavity formed by at least the amino acids Ser, His and His.
3. The composition according to claim 1 wherein said protease is a dimer.
4. A heavy atom derivative of a herpes virus protease crystal.
5. The composition according to claim 1 wherein said protease is selected from the group consisting of herpes simplex virus (HSV)-2 and HSV-1 and has an active site formed by the amino acids Ser 129, His 61, and His 148 cor-

responding to SEQ ID NO: 3 and 4.

6. The composition according to claim 1 wherein said protease is HSV-2 and said active site is characterized by the coordinates selected from the group consisting of the coordinates of Figures 2 and 3, the coordinates of Figures 8 and 9, and the coordinates of Figures 11 and 12.
7. The composition according to claim 5 wherein the protease is HSV-1, and said active site further includes amino acids Ala 131, Cys 152, Arg 156 and Arg 157 corresponding to SEQ ID NO: 3.
8. The composition according to claim 5 wherein said protease is HSV-1 and said active site is characterized by the coordinates selected from the group consisting of the coordinates of Figures 6 or 7, the coordinates of Figure 10, and the coordinates of Figure 13.
9. The composition according to claim 1 wherein said protease is human cytomegalovirus protease (CMV) and has an active site cavity formed by at least the amino acids Ser 132, His 64, and His 157 corresponding to SEQ ID NO: 1.
10. The composition according to claim 9, wherein said CMV protease active site is formed by the amino acids Ser 132, His 63, His 157, and Asp 65 corresponding to SEQ ID NO: 1.
11. The composition according to claim 9, wherein said CMV protease active site is characterized by the coordinates selected from the group consisting of the coordinates of Figure 17 or Figure 21, the coordinates of Figure 18, and the coordinates of Figure 19.
12. The composition according to claim 1, wherein said protease is varicella zoster virus (VZV) and has an active site cavity formed by at least the amino acids Ser 120, His 52, His 139, and Lys 54 corresponding to SEQ ID NO: 5.
13. The composition according to claim 12, wherein said VZV protease active site is formed by the amino acids Ser 120, His 52, His 139, Lys 54, Ser 122, Cys 143, Arg 147, and Arg 148 corresponding to SEQ ID NO: 5.
14. The composition according to claim 11, wherein said VZV protease active site is characterized by the coordinates selected from the group consisting of the coordinates of Figure 22 or Figure 23, the coordinates of Figure 24, and the coordinates of Figure 25.
15. An isolated, properly folded herpes simplex virus (HSV) 2 protease molecule, or fragment thereof, having a conformation comprising a catalytically active site formed by the interaction of three amino acids Serine, Histidine and Histidine, said active site defined by the protein coordinates of Figures 2 and 3, the distances between atoms of Figures 8 and 9, and the bond angles between interresidue atoms of Figures 11 and 12.
16. The HSV-2 protease molecule according to claim 15, wherein said molecule is a monomer characterized by a 7-stranded  $\beta$ -barrel core with seven  $\alpha$  helices, as illustrated in Figures 29A,B,C, and 31.
17. The HSV-2 protease molecule according to claim 15, wherein said molecule is a dimer characterized by the dimer interface of Fig.27 A,B, and C.
18. An isolated, properly folded herpes simplex virus (HSV) 1 protease molecule, or fragment thereof, having a conformation comprising a catalytically active site formed by the interaction of three amino acids Serine, Histidine and Histidine, said active site defined by the protein coordinates of Figure 6, the distances between atoms of Figure 10, and the bond angles between interresidue atoms of Figure 13.
19. The HSV-1 protease molecule according to claim 18, wherein said molecule is a monomer characterized by a 7-stranded  $\beta$ -barrel core with seven  $\alpha$  helices, as illustrated in Figures 30 A,B, and 31.
20. The HSV-1 protease molecule according to claim 18 wherein said molecule is a dimer characterized by the dimer interface of Fig.28 A, and B.
21. An isolated, properly folded cytomegalovirus (CMV) protease molecule, or fragment thereof, having a conformation comprising a catalytically active site formed by the interaction of three amino acids Serine, Histidine and Histidine, said active site defined by the protein coordinates of Figure 17, the distances between atoms of Figure 18, and the

bond angles between interresidue atoms of Figure 19.

22. The CMV protease molecule according to claim 21, wherein said molecule is a monomer characterized by a 7-stranded  $\beta$ -barrel core with seven  $\alpha$  helices, as illustrated in Figures 32 A, 32B, and 34.
23. The CMV protease molecule according to claim 21, wherein said molecule is a dimer characterized by the dimer interface of Fig. 36.
24. An isolated, properly folded varicella zoster virus (VZV) protease molecule, or fragment thereof, having a conformation comprising a catalytically active site formed by the interaction of three amino acids Serine, Histidine and Histidine, said active site defined by the protein coordinates of Figure 22, the distances between atoms of Figure 24, and the bond angles between interresidue atoms of Figure 25.
25. The VZV protease molecule according to claim 24 wherein said molecule is a monomer characterized by a 7-stranded  $\beta$ -barrel core with eight  $\alpha$  helices, as illustrated in Figures 33A, 33B, and 35.
26. The VZV protease molecule according to claim 24 wherein said molecule is a dimer characterized by the dimer interface of Fig. 37.
27. A peptide, peptidomimetic, synthetic or natural product molecule which binds with the active site cavity of a herpes virus protease composition, derivative or molecule of any of claims 1 to 26.
28. A method of identifying an inhibitor compound capable of binding to, and inhibiting the proteolytic activity of, a herpes protease, or any other protease characterized by the Ser-His-His catalytic triad, said method comprising:
  - introducing into a suitable computer program information defining an active site conformation of a herpes protease molecule comprising a catalytically active site formed by at least the interaction of three amino acids Serine, Histidine and Histidine, wherein said program displays the three-dimensional structure thereof;
  - creating a three dimensional representation of the active site cavity in said computer program;
  - displaying and superimposing the model of said test compound on the model of said active site;
  - assessing whether said test compound model fits spatially into the active site;
  - incorporating said test compound in a biological protease activity assay for a protease characterized by said active site; and
  - determining whether said test compound inhibits proteolytic activity in said assay.
29. The method according to claim 28, wherein the protease is liganded or unliganded herpes simplex virus (HSV)2 according to any of claims 15 to 17.
30. The method according to claim 28, wherein the protease is herpes simplex virus (HSV) 1 according to any of claims 18 to 20.
31. The method according to claim 28, wherein the protease is cytomegalovirus according to any of claims 21 to 23.
32. The method according to claim 28, wherein the protease is varicella zoster virus according to any of claims 24 to 27.
33. A peptide, peptidomimetic, synthetic or natural product molecule identified by the method of claim 28.
34. A method for solving a crystal form comprising using the structural coordinates of herpes protease crystal or portions thereof, to solve a crystal form of a mutant, homologue or co-complex of said protease by molecular rearrangement.
35. A method of drug design comprising using the structural coordinates of a herpes protease crystal to computationally evaluate a chemical entity for associating with the active site of a herpes virus protease.
36. The method according to claim 35, wherein said entity is a competitive, non-competitive or uncompetitive inhibitor, binds to or inhibits the proteolytic activity of a herpes virus protease.
37. The method of drug design comprising using the structure coordinates of a herpes virus protease to identify an

intermediate in a chemical reaction between said protease and a compound which is a substrate or inhibitor of said protease.

38. The method according to claims 25 or 37, wherein said structure coordinates are selected from the group consisting of the coordinates of Figures 2, 3, 6, 8-20, 22 and 22-26.

39. A method for identifying inhibitors which competitively bind to the dimeric interface of a herpes virus protease molecule or fragment thereof, which protease is characterized by a catalytically active site formed by the interaction of three amino acids Serine, Histidine and Histidine, said method comprising the steps of:

providing the coordinates of said active site, and dimeric interface of said protease to a computerized modeling system;  
identifying compounds which will bind to, or interfere with, the dimeric interface; and  
screening the compounds identified for protease inhibitory bioactivity.

40. A method for identifying inhibitors which competitively bind to the active site of a herpes virus protease molecule or fragment thereof characterized by a catalytically active site formed by the interaction of three amino acids Serine, Histidine and Histidine, said method comprising the steps of:

providing the coordinates of said active site of the protease to a computerized modeling system;  
identifying compounds which will bind to the structure; and  
screening the compounds identified for protease inhibitory bioactivity.

41. The method according to claim 40, wherein the protease is liganded or unliganded herpes simplex virus (HSV) 2 according to any of claims 15 to 17.

42. The method according to claim 40, wherein the protease is herpes simplex virus (HSV) 1 according to any of Claims 18 to 20.

43. The method according to claim 40, wherein the protease is cytomegalovirus (CMV) according to any of claims 21 to 23.

44. The method according to claim 40, wherein the protease is varicella zoster virus (VZV) according to any of claims 24 to 26.

45. A modified varicella zoster virus (VZV) protease selected from the group consisting of the sequence of Fig. 45A SEQ ID NO: 7, the sequence of Fig. 45B SEQ ID NO: 8, the sequence of Fig. 45C SEQ ID NO: 9, the sequence of Fig. 45D SEQ ID NO: 10, and the sequence of Fig. 45E SEQ ID NO: 11.

46. A method of forming a crystal of VSV protease comprising crystallizing a modified VSV protease of claim 45.

47. The method according to claim 46, wherein said VSV protease has the sequence of Fig. 45E SEQ ID NO: 11.

48. In a bioassay for identifying inhibitors of VSV protease, wherein said bioassay comprises the step of exposing a VZV protease to a candidate inhibitor, the improvement comprising using as said VSV protease a protease of claim 45.

49. A method for purifying a varicella zoster virus (VZV) protease with an intact mature carboxy terminal sequence of LQA, said method comprising purifying an authentic VZV sequence with an amino acid sequence interposed between said LQA terminal amino acids and a hexahistidine sequence on an NiNTA column.

50. The method according to claim 49, wherein said VZV sequence is that of Fig. 45E SEQ ID NO: 11.

51. A method for purifying a cytomegalovirus (CMV) protease with an intact mature carboxy terminal sequence of LQA, said method comprising purifying an authentic CMV sequence with an amino acid sequence interposed between said LQA terminal amino acids and a hexahistidine sequence on an NiNTA column.

52. A method for purifying a herpes simplex virus (HSV) 1 protease with an intact mature carboxy terminal sequence of LQA, said method comprising purifying an authentic HSV1 sequence with an amino acid sequence interposed

between said LQA terminal amino acids and a hexahistidine sequence on an NiNTA column.

53. A method for purifying a herpes simplex virus (HSV) 2 protease with an intact mature carboxy terminal sequence of LQA, said method comprising purifying an authentic HSV2 sequence with an amino acid sequence interposed  
5 between said LQA terminal amino acids and a hexahistidine sequence on an NiNTA column.
54. A computer readable medium having stored thereon a model of the crystal structure of the catalytic site domain of a herpes protease.
- 10 55. A protease crystal structure characterized by a catalytic site triad comprising Serine, Histidine and Histidine.
56. The protease crystal structure according to claim 55, wherein said structure is a truncated herpes protease.

15

20

25

30

35

40

45

50

55

Fig. 26

	Phi $C_n-1-N_n-C_{An}-C$	Psi $N-CA-C-N_{n+1}$	Omega 1 $CA_n-1-C_n-1-N_n-CA_n$	Omega 2 $CA_n-C_n-N_{n+1}-CA_{n+1}$	Chi1 $N-CA-CB-CG$	Chi2 $CA-CD-CG-XD$
Ser120	120	-120	118	-179	-180	-141
His52	71	13	-178	179	-69	-81
His139	-158	155	178	180	74	111

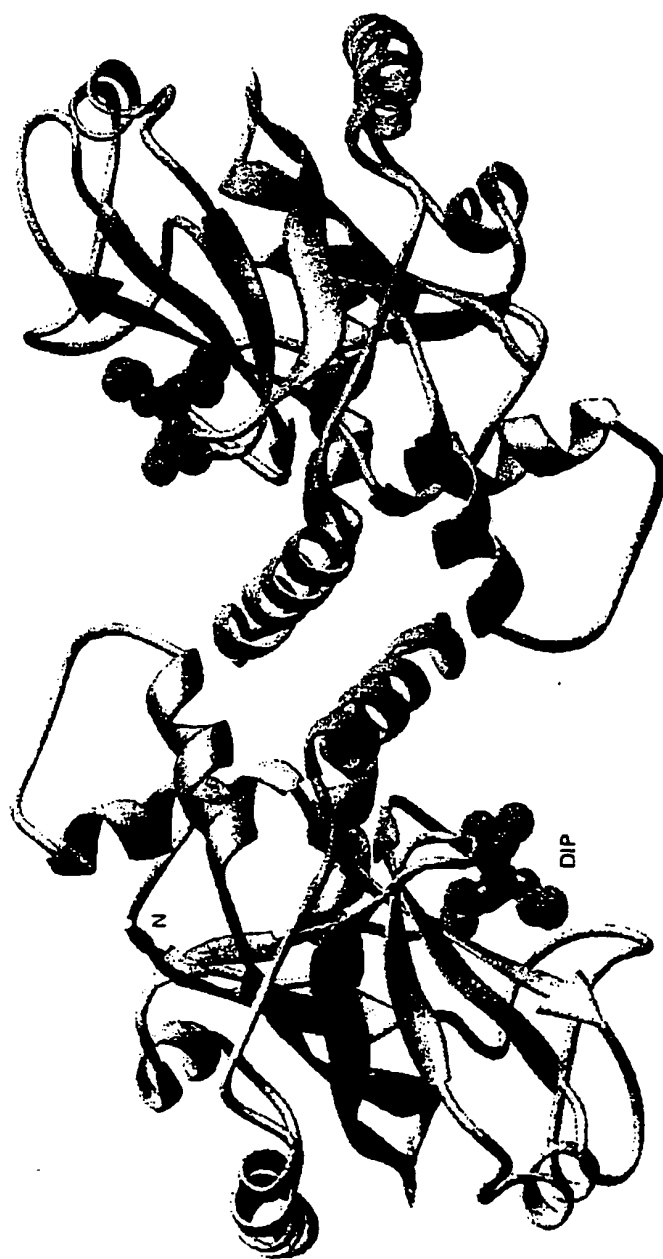


Fig. 27A



Fig. 27B



Fig. 27C



Fig. 28A

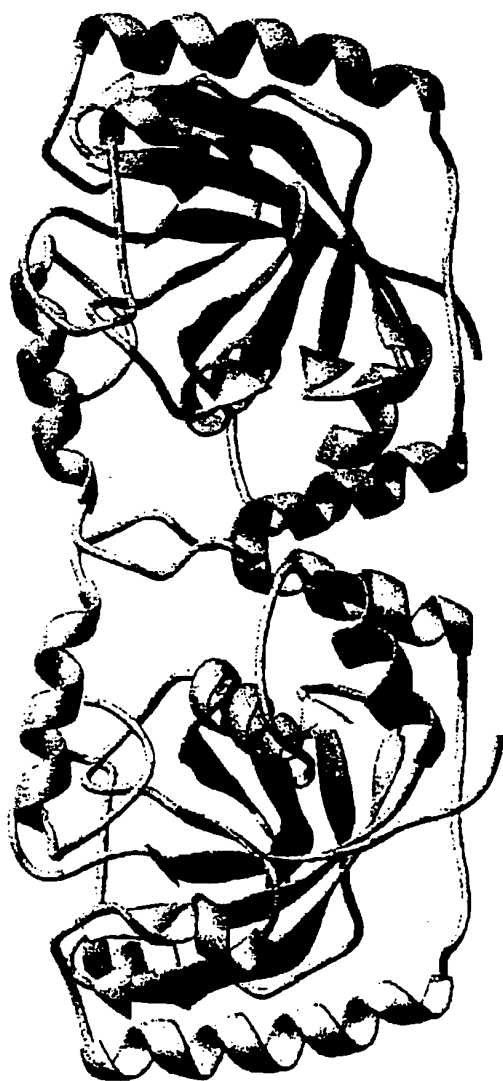


Fig. 28B

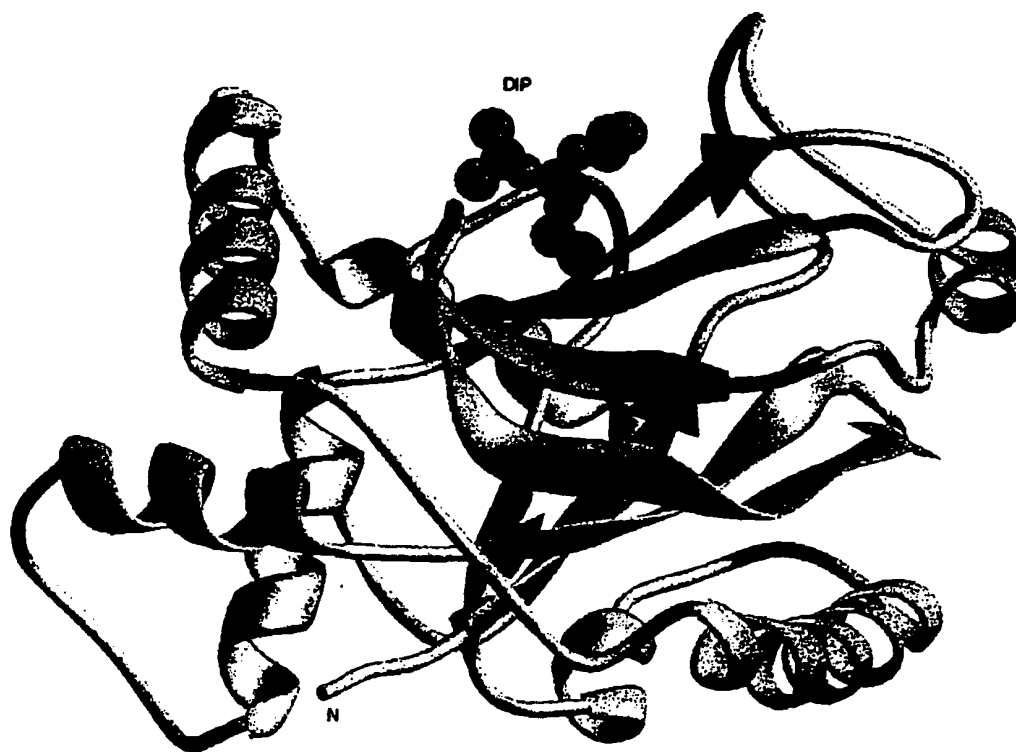


Fig. 29A

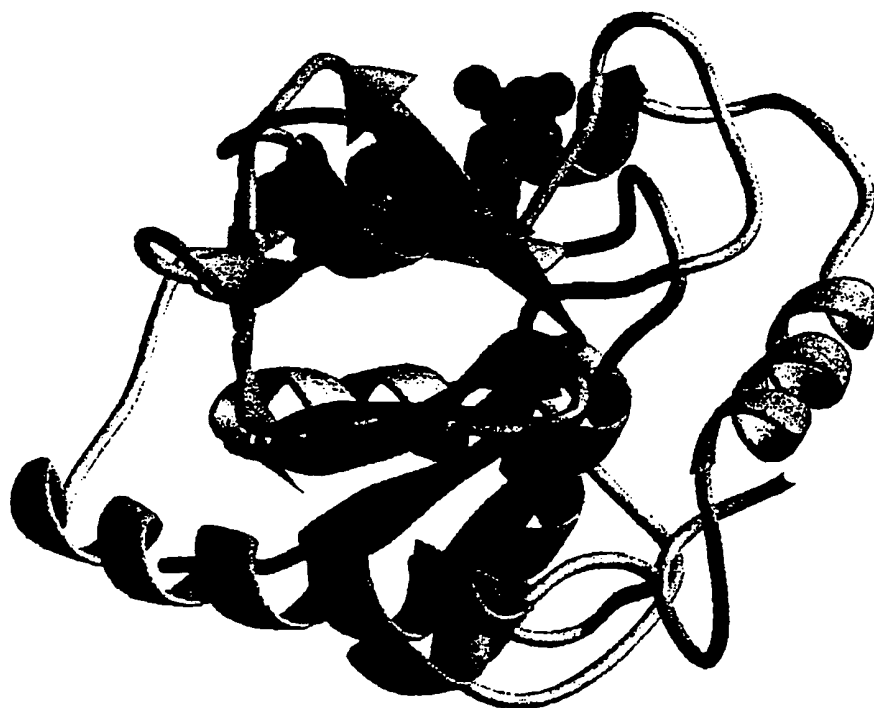


Fig. 29B

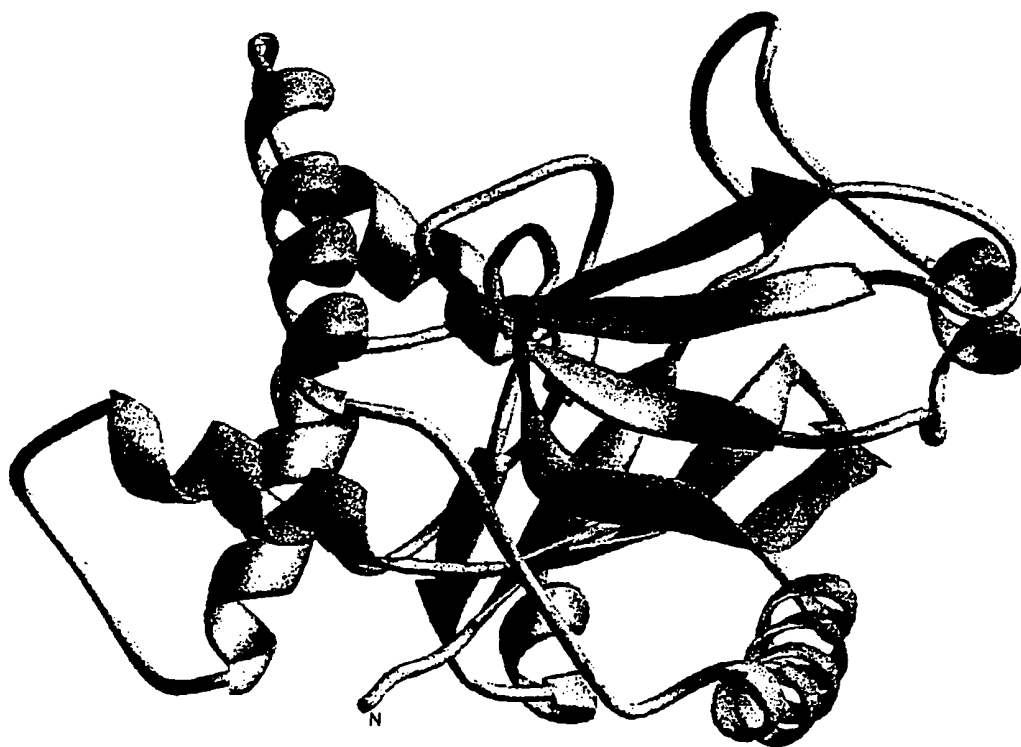


Fig. 29C



Fig. 30A

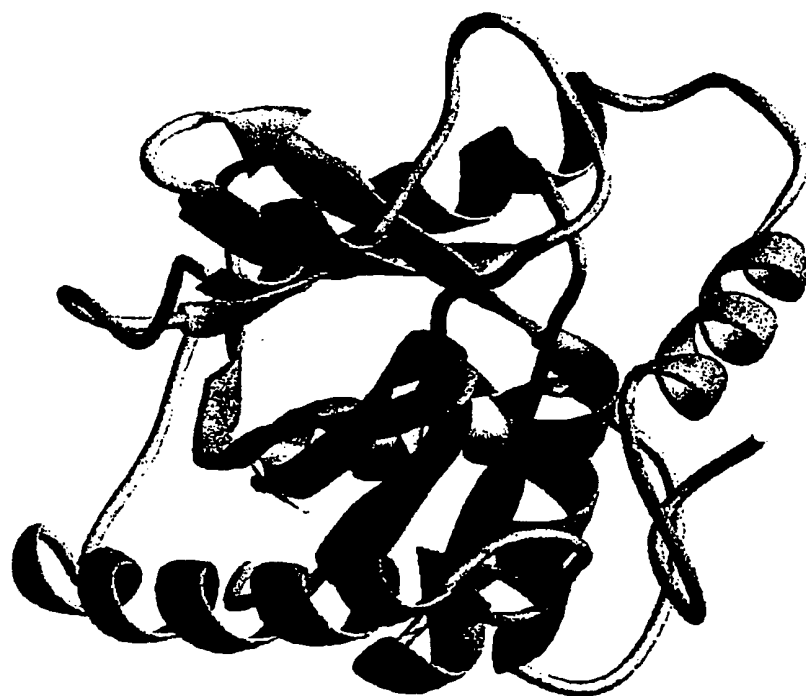


Fig. 303

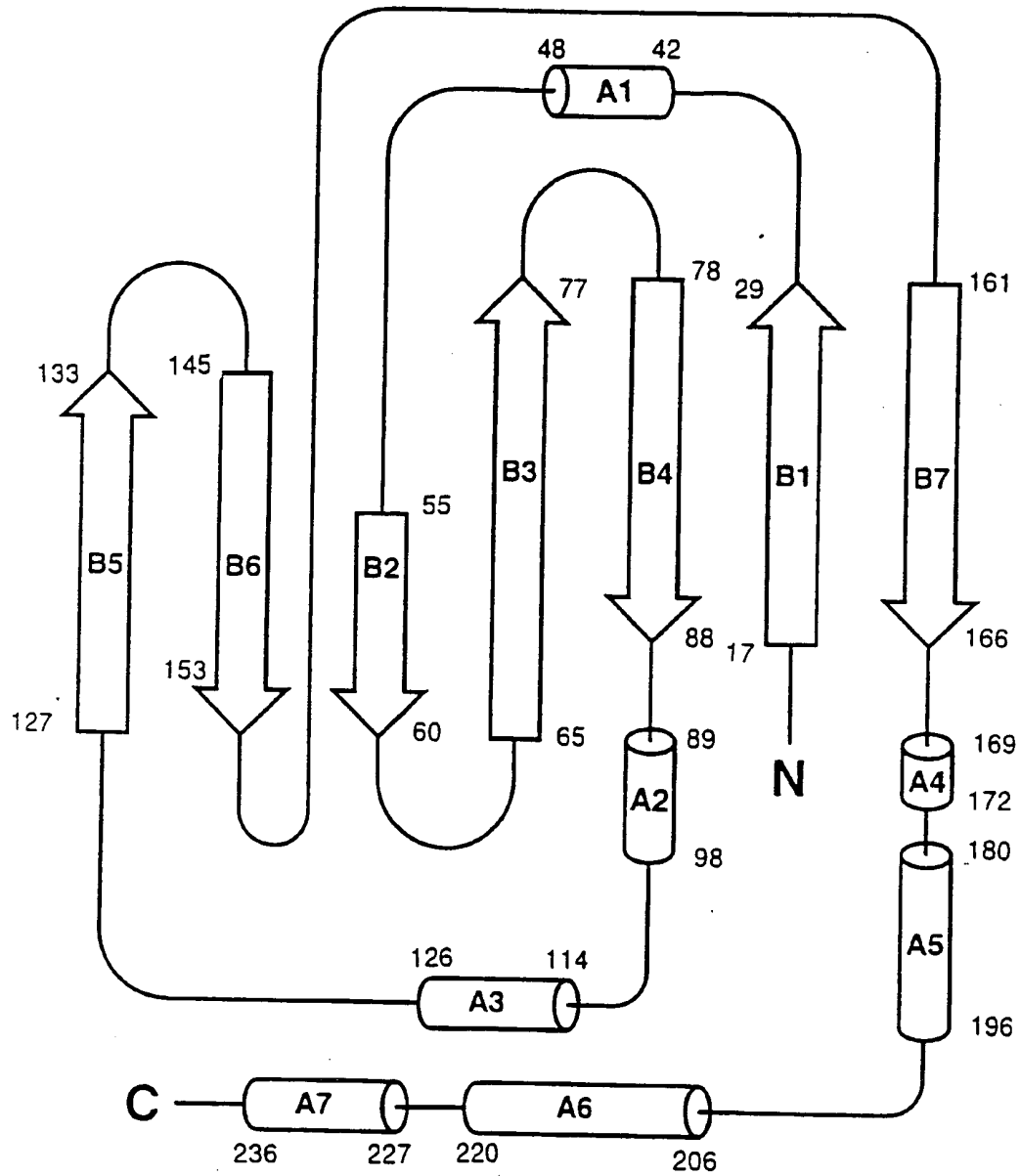


Fig. 31

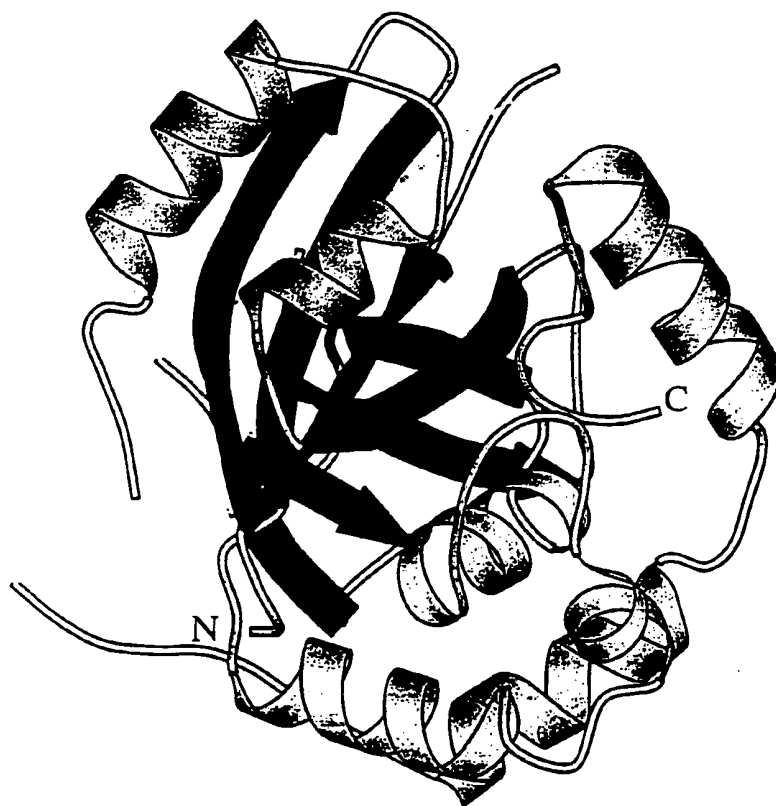


Fig. 32A

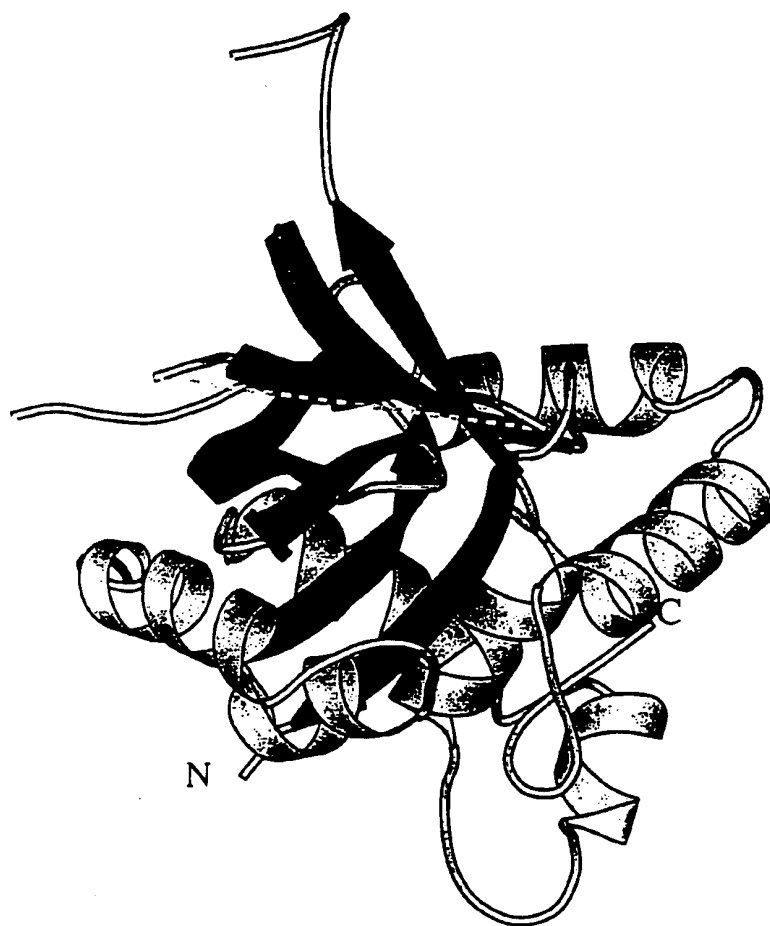


Fig. 32B

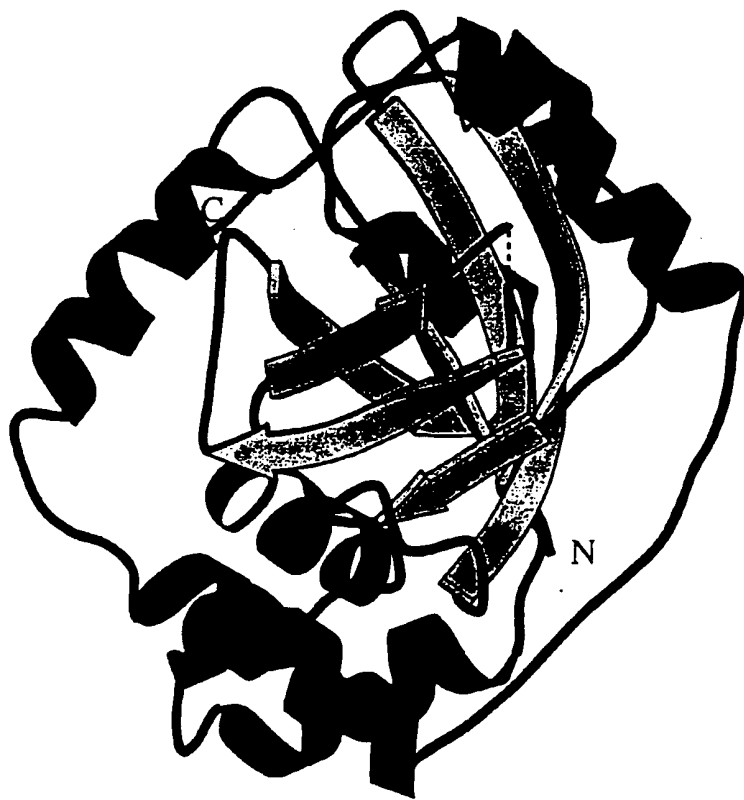


Fig. 33A



Fig. 338

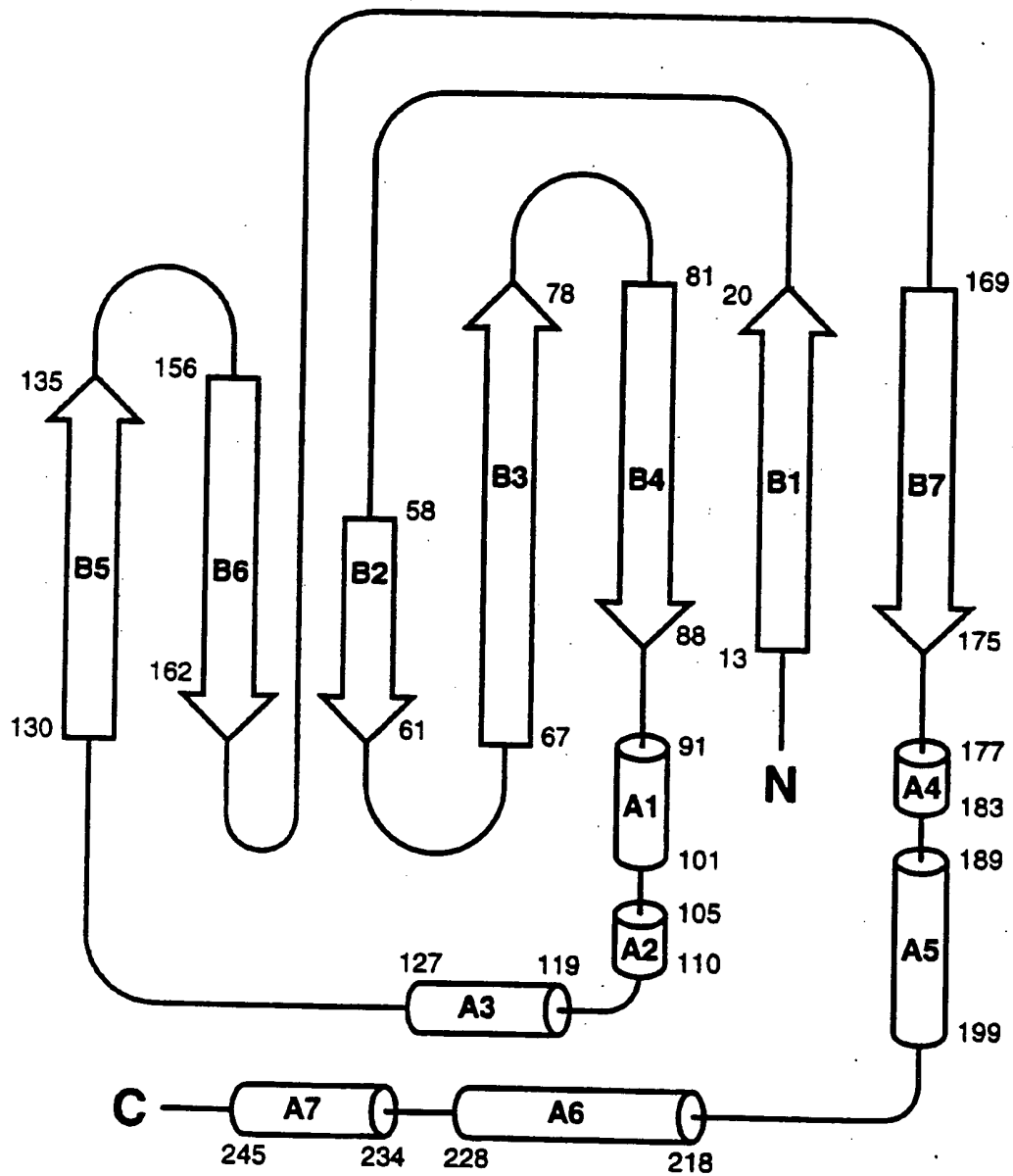


Fig. 34

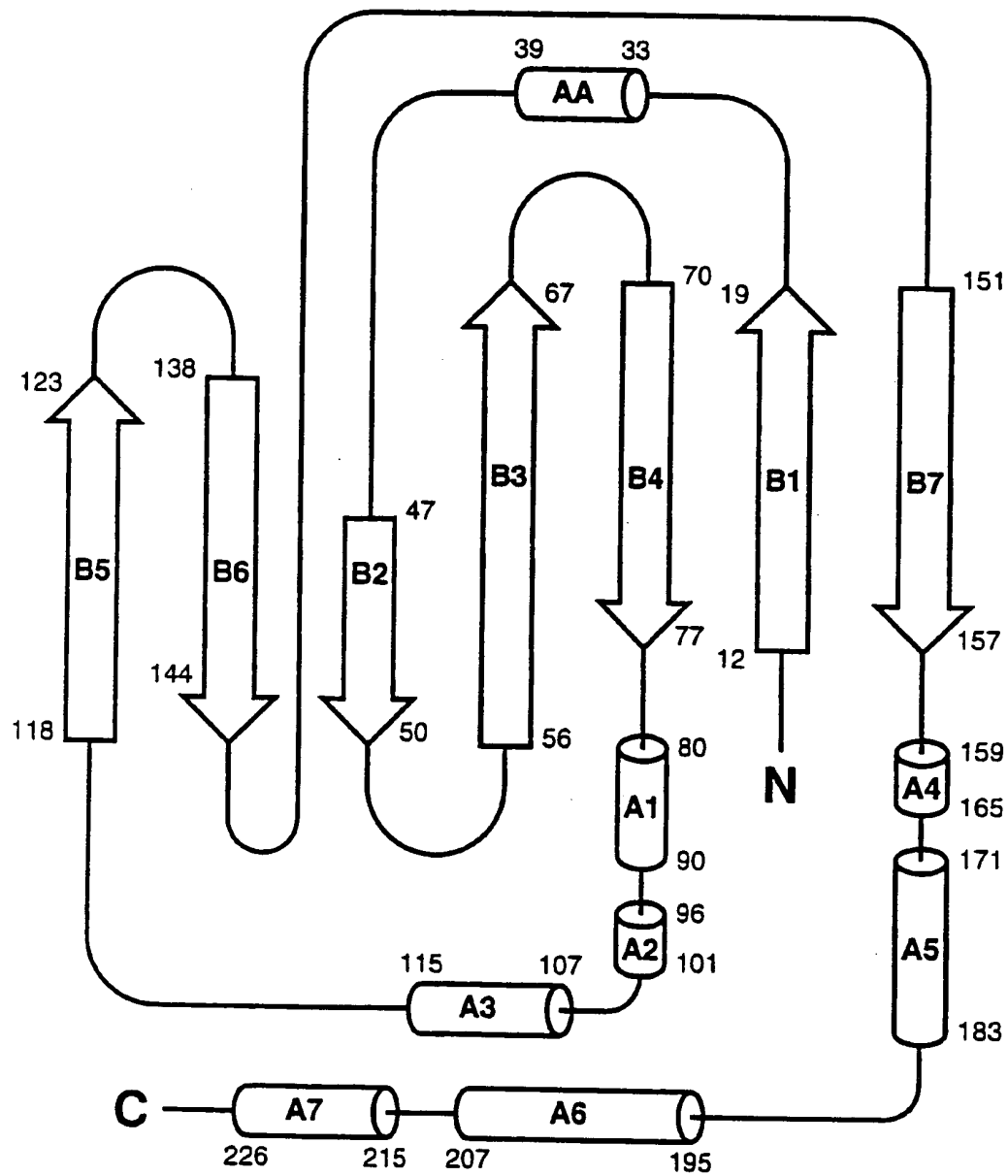


Fig. 35

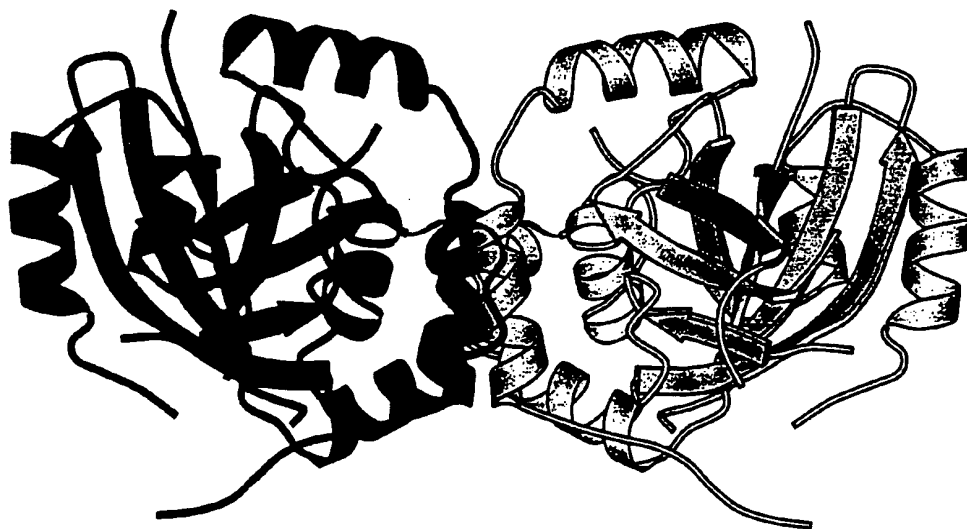


Fig. 36A

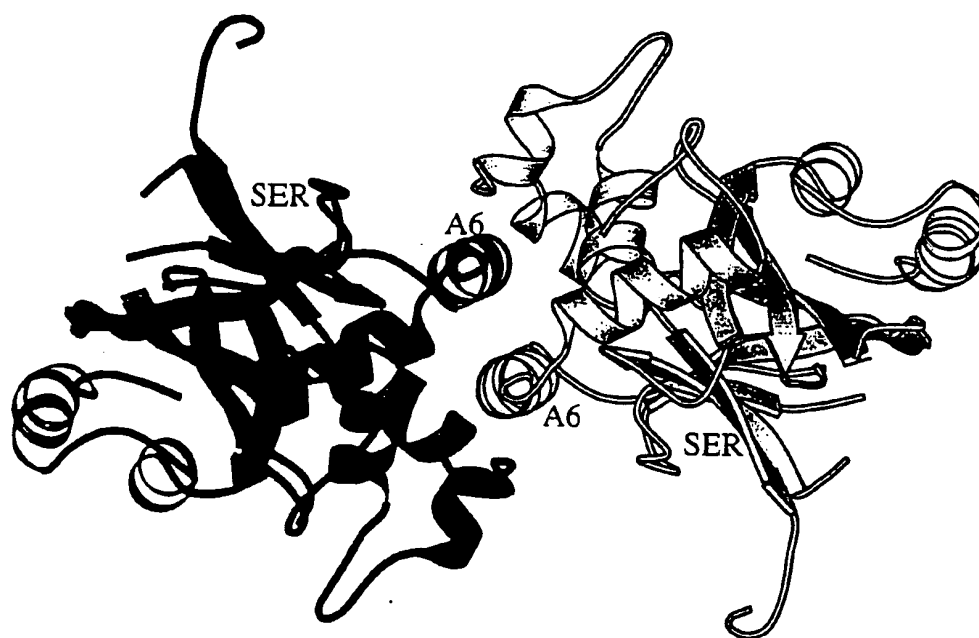


Fig. 36B



Fig. 37

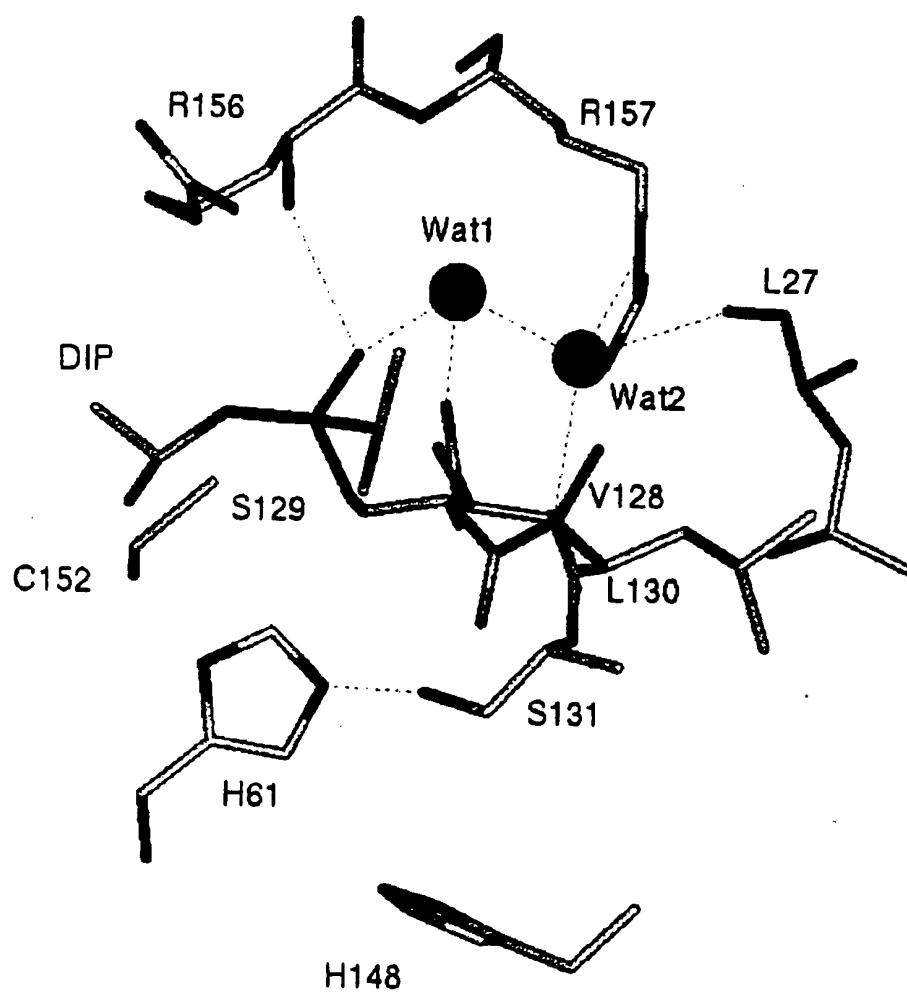
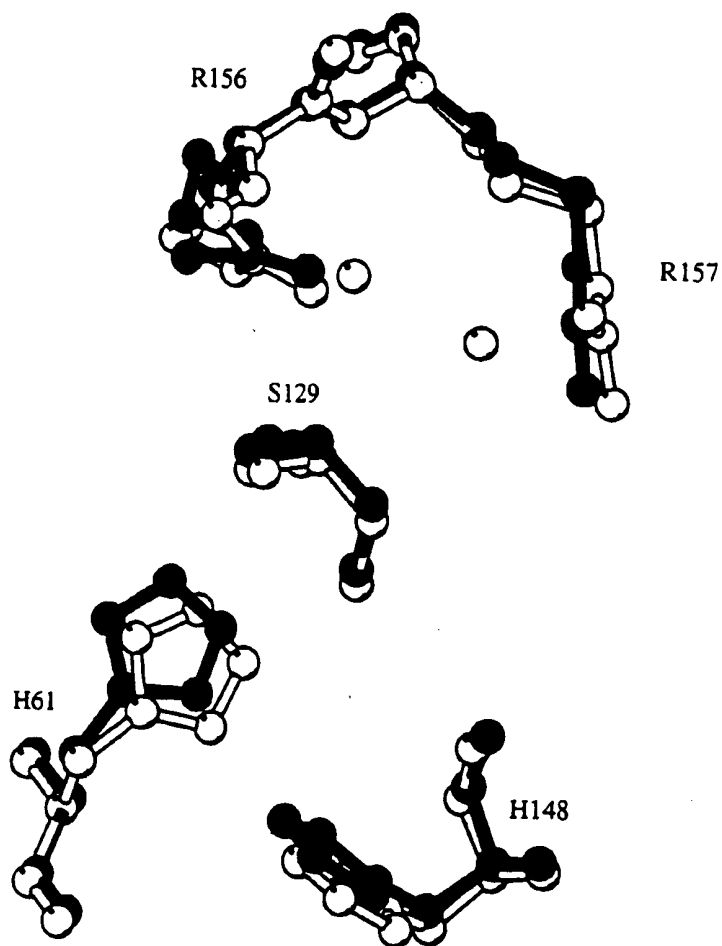


Fig. 38A

Fig. 388



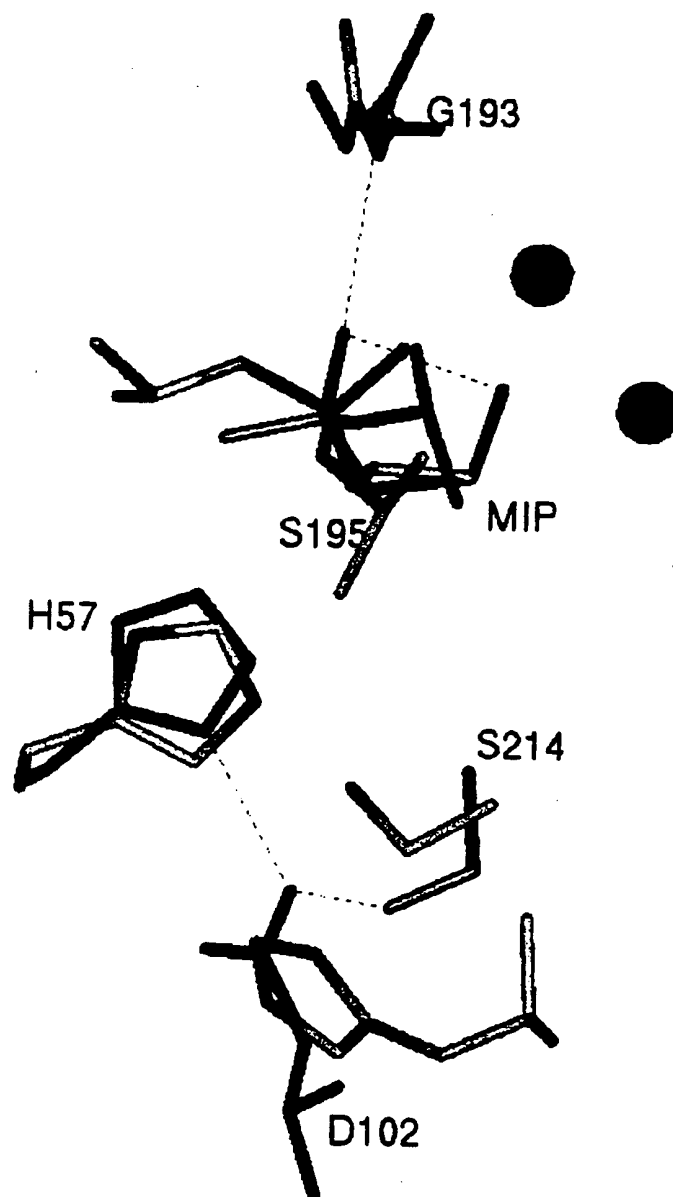


Fig. 38C

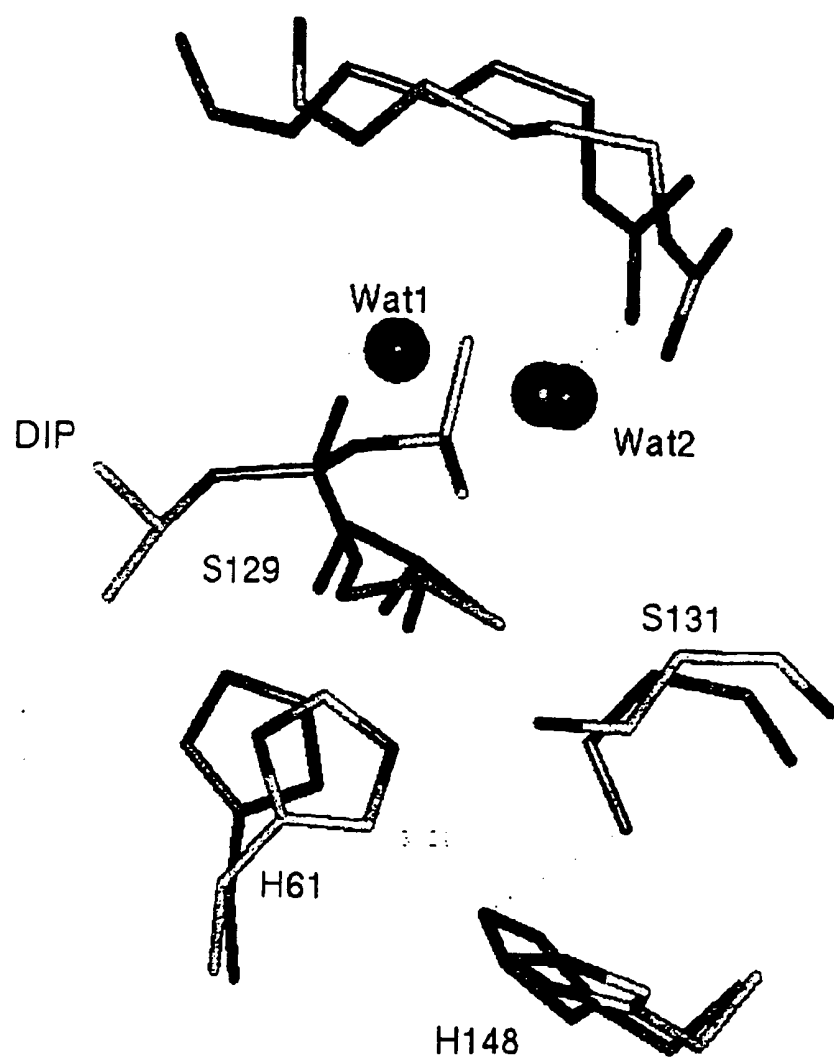


Fig. 38D

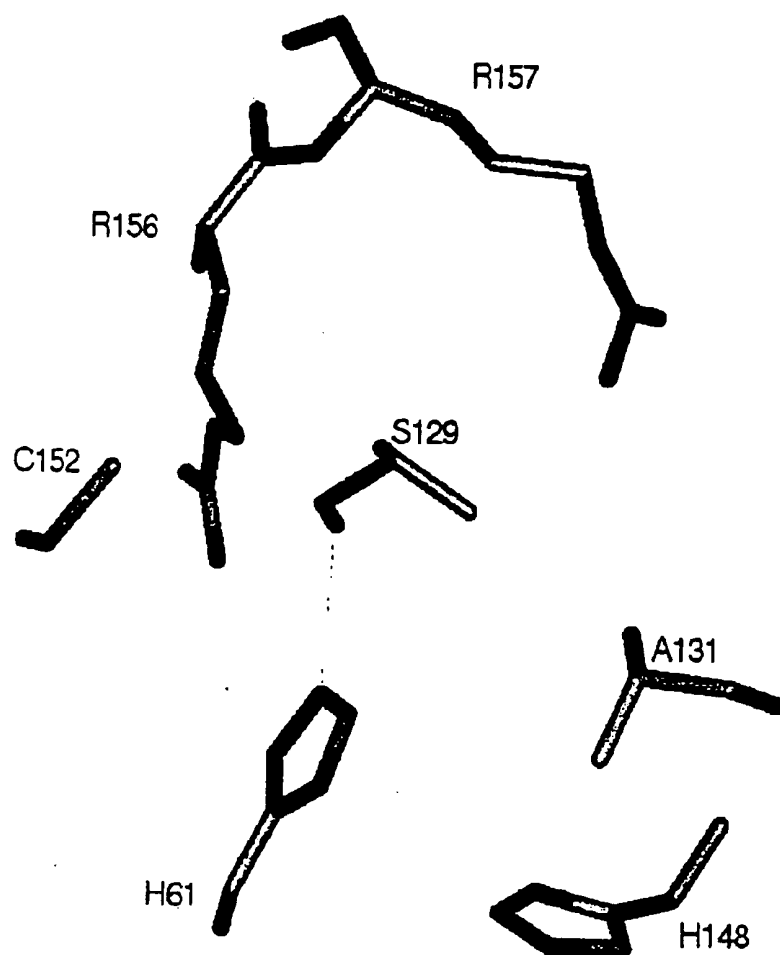


Fig. 39A

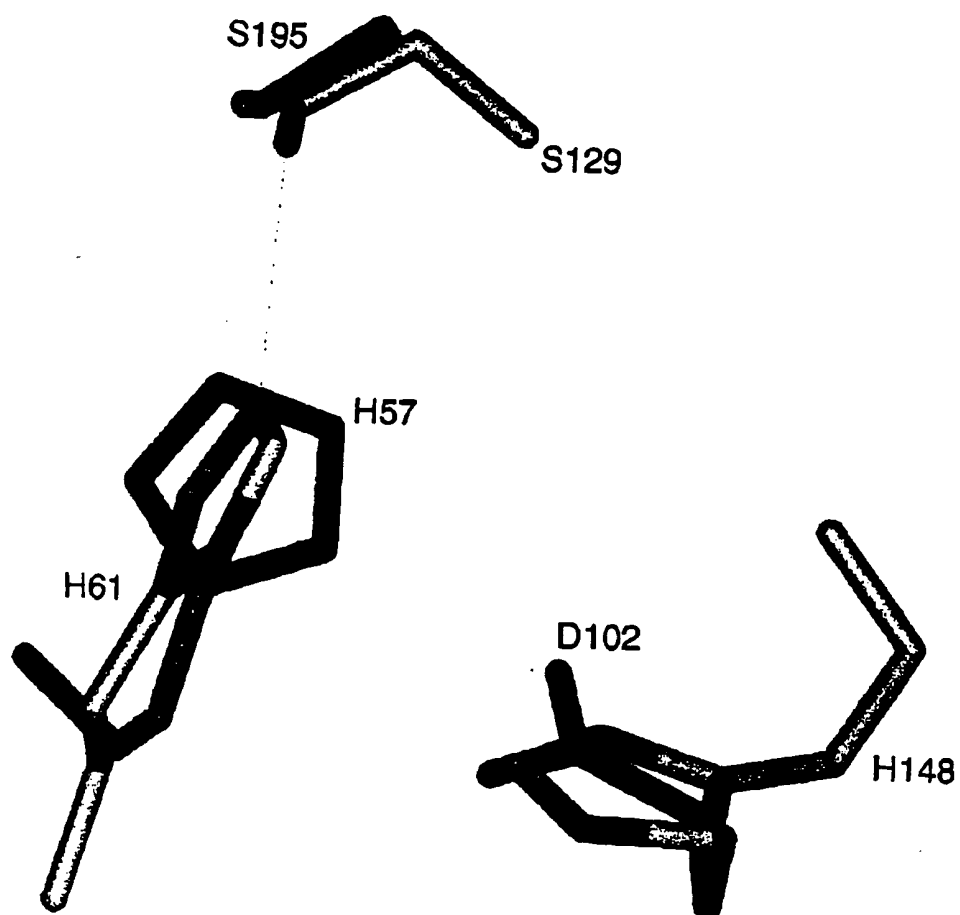


Fig. 398

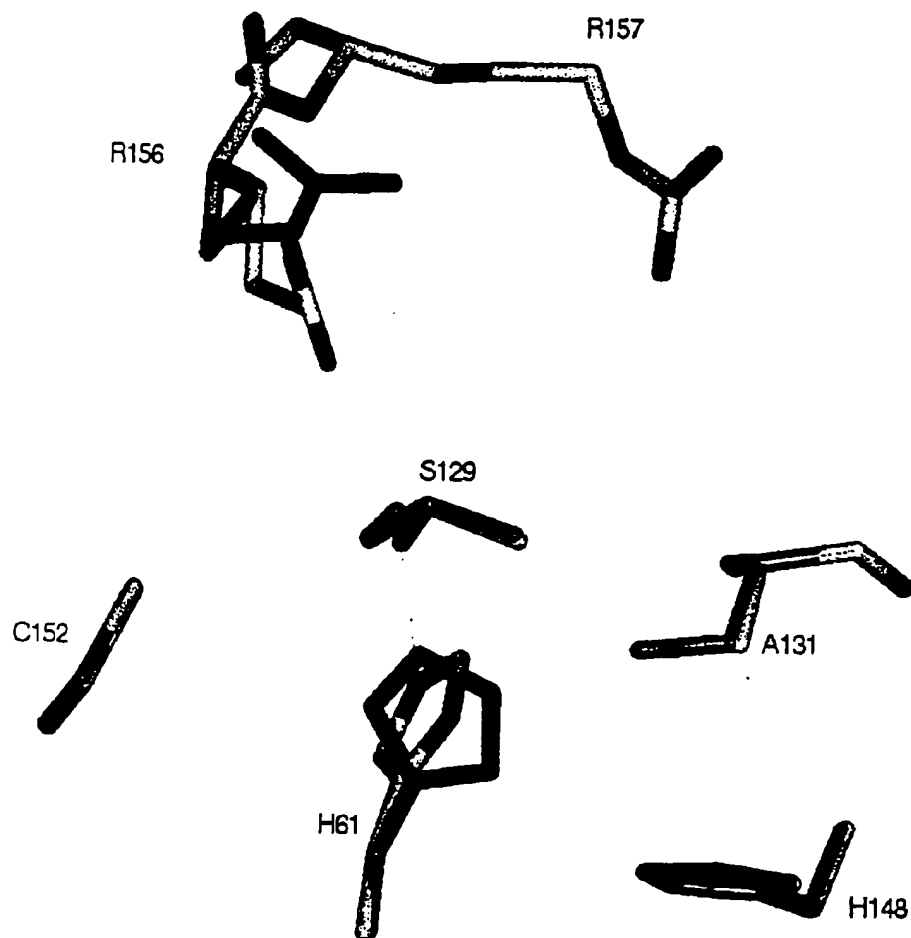


Fig. 39C

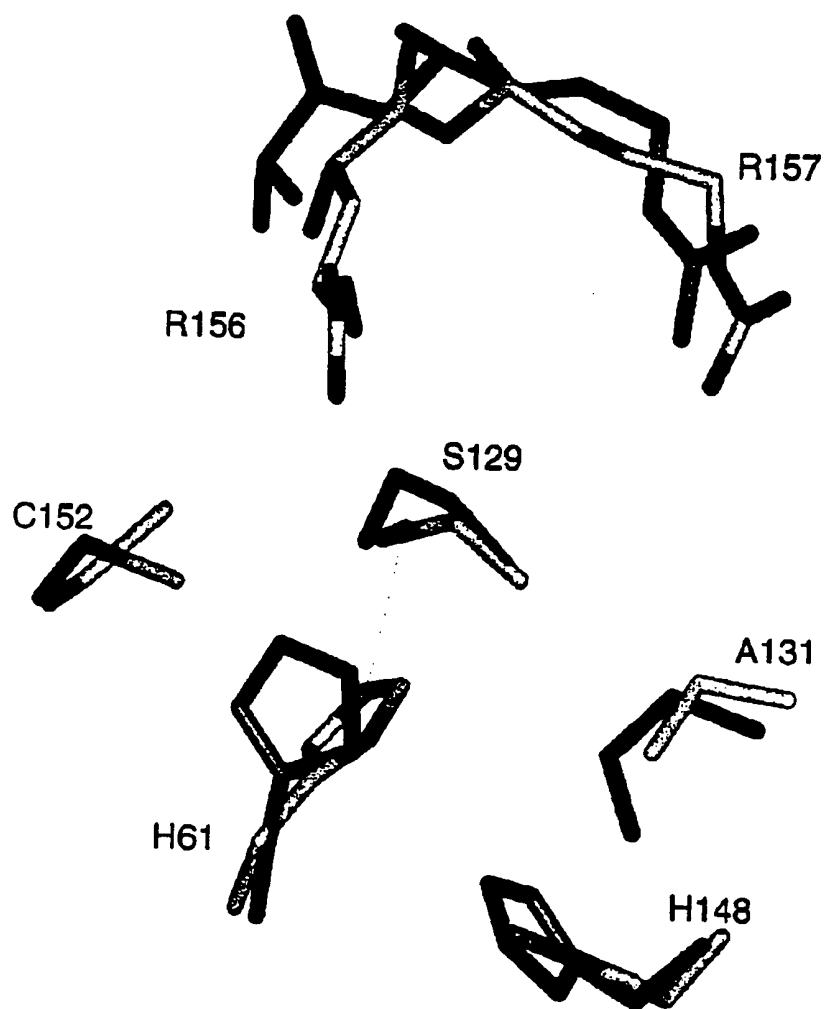
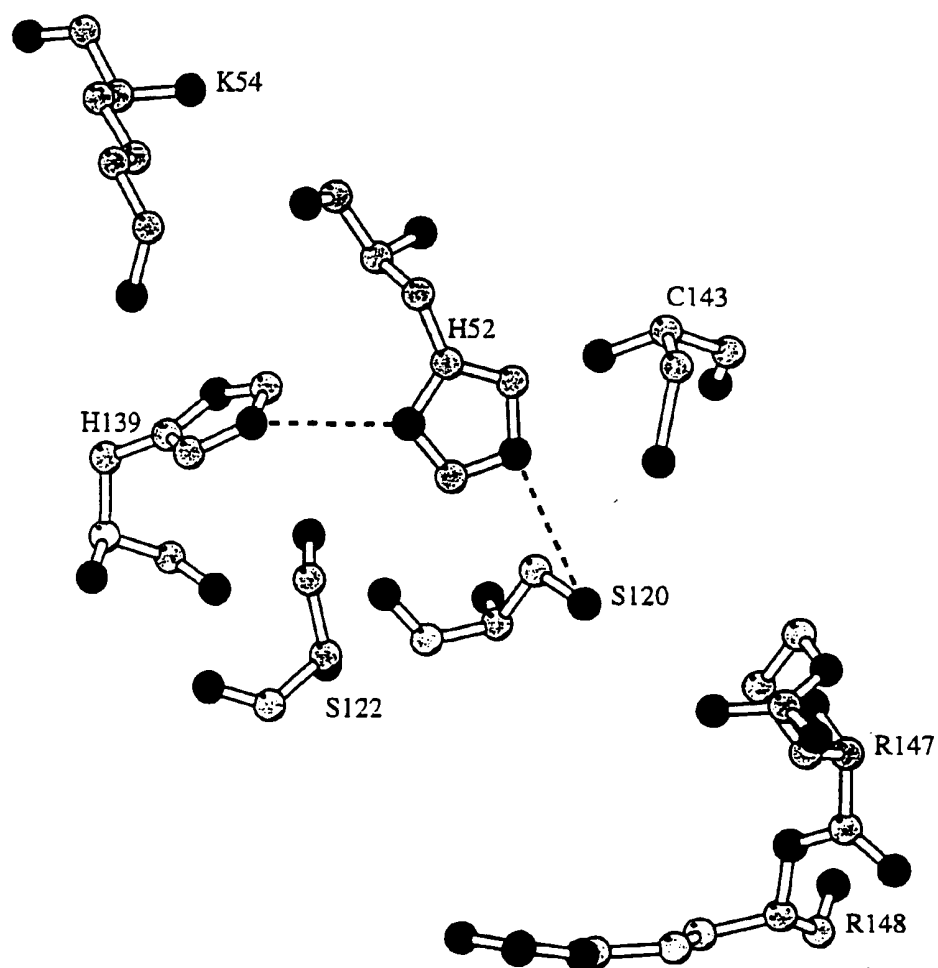
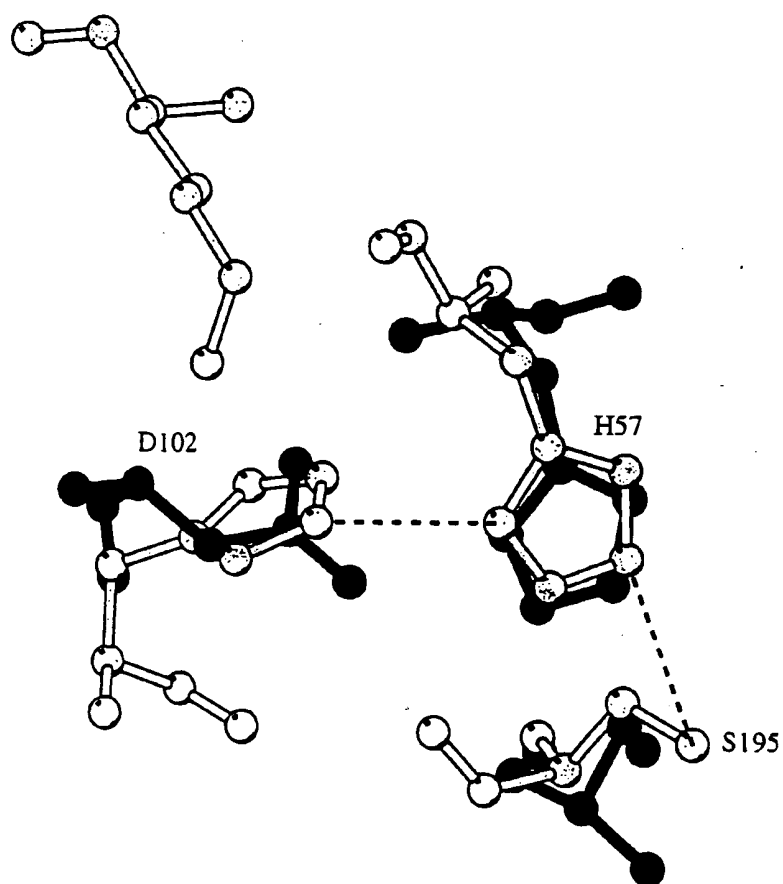


Fig. 39D



VZVP Active Site

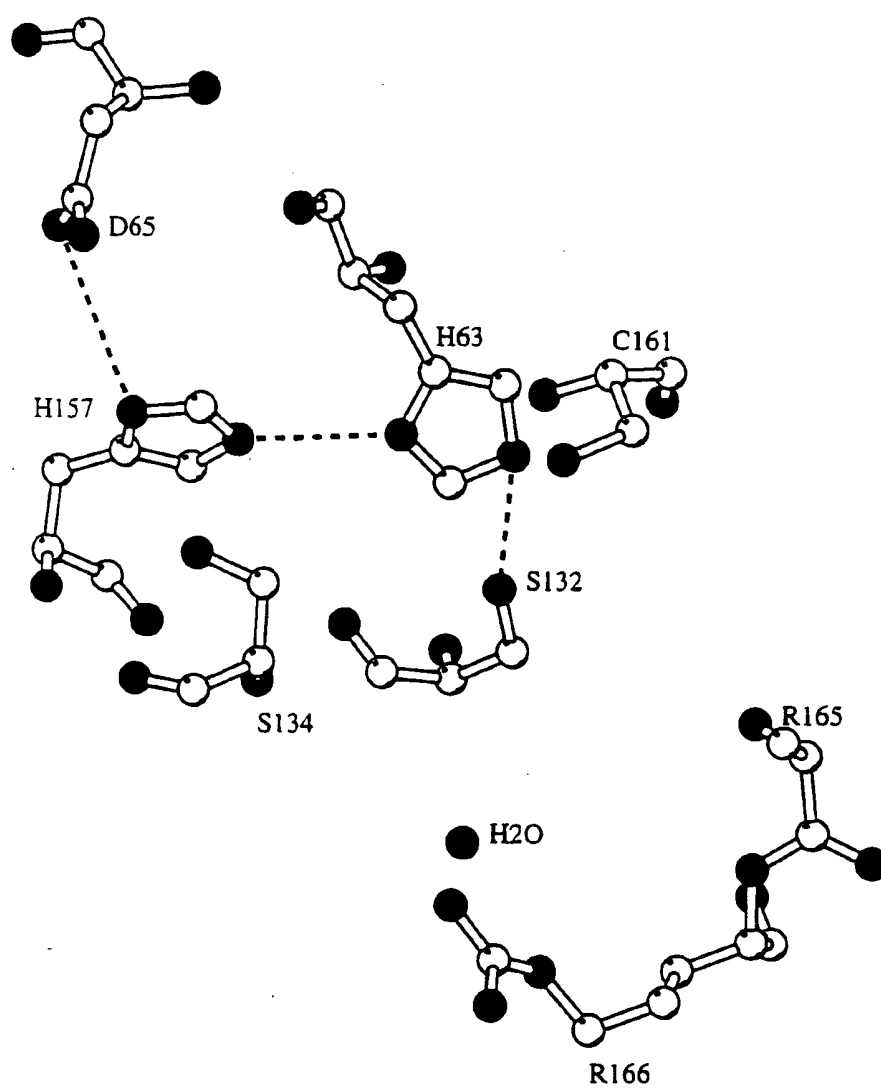
Fig. 40A



Light-VZV, Dark-Trypsin

Fig. 408

Fig. 41A



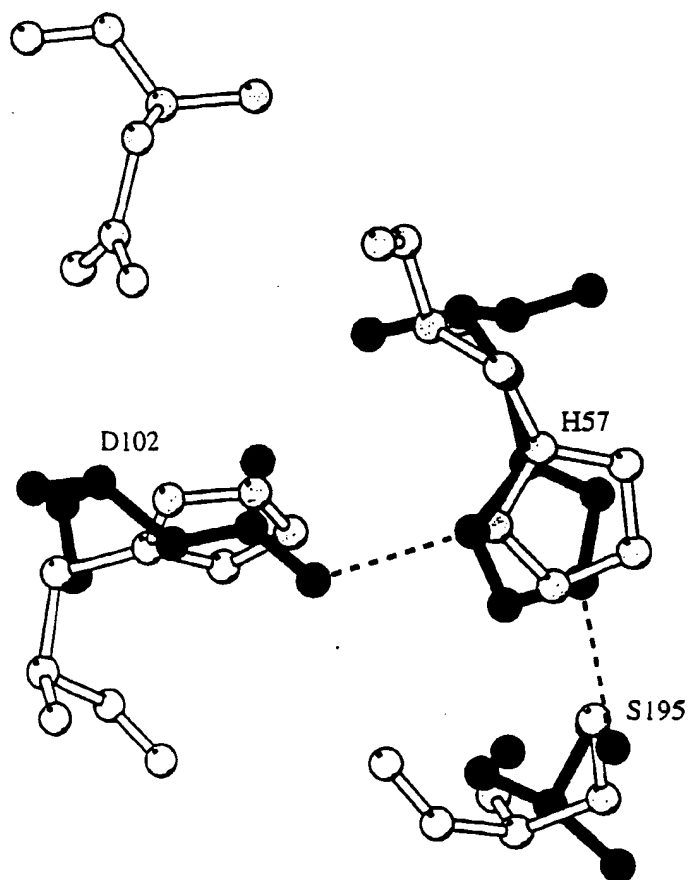


Fig. 418

Figure 42. Structure Determination Statistics

Data Set	Resol'n (Å)	Observed	Unique	Compleat (%)	R <sub>sym</sub>	R <sub>iso</sub>	# sites	Phasing Power	R <sub>collis</sub>
Native	2.5	78502	15788	90.8	0.095	---	---	---	---
KAuCN	3.0	31511	10071	98.2	0.078	15.9	2	1.14	0.74
LuCl <sub>3</sub>	3.0	31104	9964	97.3	0.078	12.4	1	1.51	0.79
PrCl <sub>3</sub>	3.0	31645	10179	99.0	0.090	13.6	1	1.4	0.82
YbSO <sub>4</sub>	4.0	10495	3787	78.7	0.088	14.6	1	2.01	0.72
GdCl <sub>3</sub>	4.0	18308	4268	96.6	0.166	15.7	1	1.56	0.79
SmCl <sub>3</sub>	4.0	25172	4327	97.6	0.114	8.8	1	1.47	0.86

MIR Overall Mean Figure of Merit (15-3.0 Å): 0.62 Å

Overall Figure of Merit After Phase combination: 0.67 Å

Mean Figure of Merit following density modification 100-3.0Å: 0.838 Å

**X-PLOR refinement: 10-2.5 Å**

No. Reflections Used (F&gt;2σ): 14605

R-factor: 20.5%

No. Protein atoms (non-H): 3364

No. Solvent Atoms: 43

Rms bond length: 0.016 Å

Rms bond angles: 1.919°

Fig. 43

Data Set	Resolu'n (Å)	Observed (>1σ)	Unique (>1σ)	Complete (%)	R <sub>m</sub> (%)	R <sub>iso</sub> (%)	# of sites	# of Phasing	R <sub>cullis</sub>	R <sub>c</sub> (ano)
Native	2.5	22880	7644	90	6.0	-----	--	---	---	--
MeHgCl	3.2	17566	3932	97	10.1	28.0	5	2.8	0.50	0.92
MeHgCl	3.5	3698	1958	76	19.8	36.5	3	1.4	0.75	--
UO <sub>2</sub> Ac <sub>2</sub>	3.8	9122	2330	99	23.2	27.5	2	1.1	0.83	--
BakerHg <sub>2</sub>	3.2	14867	4425	97	9.8	30.7	4	1.9	0.68	0.97
LuCl <sub>3</sub>	3.2	9984	3734	95	10.8	23.4	3	2.0	0.63	--
K <sub>2</sub> PtCl <sub>4</sub>	4.1	3017	1673	79	13.1	25.9	2	1.4	0.77	--
SmCl <sub>3</sub>	3.7	12708	2804	98	13.7	26.3	3	2.4	0.54	0.97

MIRAS Figure of Merit (30 - 3.2 Å): 0.70

## XPLOR Refinement:

Resolution included	7.0 - 2.5 Å	# of solvent atoms (non-H)	73
# of reflections used (>1σ)	7193	Mean coordinates error	0.40 Å
R-factor	0.185	Rms bond length	0.017 Å
# of protein atoms (non-H)	1504	Rms bond angle	2.2 degree
(202 amino acids)			



Light-CMV, Dark-VZV, RMSCA 1.3A

Fig. 44